Abstract

Error Mitigation and Detection in Circuit Quantum Electrodynamics Powered by QND Measurements

Jacob Charles Curtis

2024

Superconducting cavities coupled to transmon ancillae are a promising platform for fault-tolerant quantum computation. This pairing produces universal control over a long-lived bosonic mode, in whose many levels we encode a logical qubit or physical quantum system, such as the vibrational modes of a molecule. However, unwanted static interactions and spurious couplings to the environment limit the accuracy of computations. In this dissertation, we tackle each of these error sources in turn. Our solutions exploit the ability of the transmon to perform high-fidelity measurements of the cavity mode without disturbing its stored information. First, we present our efforts to reduce the residual cavity non-linearity, which causes dephasing in many quantum codes. We discuss multiple methods to measure the non-linearity and show that our mitigation technique can fully cancel cavity self-Kerr. Next, we construct a new photon-number resolving measurement that extracts multiple bits of information from the cavity in a single shot using multiple transmon measurements. The errors that occur in this measurement have a predictable form and can be inverted to reduce ensemble error by an order of magnitude. Finally, we develop and demonstrate a gate that detects errors in real time. The resources required for fault-tolerance are high, but can be relaxed if the goal is only to detect, but not correct errors. Our error-detected gate reaches a fidelity of 0.9995 driven by improvements in numerical gate optimization and transmon three level control.

Error Mitigation and Detection in Circuit Quantum Electrodynamics

Powered by QND Measurements

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Jacob Charles Curtis

Dissertation Director: Robert J. Schoelkopf

May 2024

Copyright © 2024 by Jacob Charles Curtis

All rights reserved.

To William P. Curtis,

in loving memory.

Contents

List of Figures

List of Tables

Acknowledgements

No Ph.D. is completed alone, and I am forever grateful for the help and support I have received at every stage of my physics education. It began at the University of Michigan, where I was pushed in my math and physics classes by my professors and friends. I have many fond memories of doing problem sets in West Hall classrooms while watching college football (Go Blue!). My interest in quantum physics soon formed and grew during the time I spent in ALEX KUZMICH's lab working with JACOB LAMPEN on atomic clocks and Rydberg states of Rb. I am grateful to both for their research mentorship and graduate school advice.

Next, I joined ROBERT SCHOELKOPF's lab here at Yale. Rob has equipped his group with the resources and personnel to churn out top-tier research for more than two decades. His success at hiring outstanding post-docs and students has perpetuated that standard and produced an environment where the answer to any question requires only a quick survey of the 4th floor of Becton. Rob's mastery of counting arguments to estimate the precision of a measurement before it's performed is remarkable, and has certainly saved me from pursuing multiple dead ends. His quickness at identifying the next useful experiment is only rivaled by his speed with a Monty Python reference. I'm excited to see where Rob leads our field and thankful for the guidance and scientific freedom he gave me over the last six years.

My dissertation committee includes two of our theory colleagues from across the street: STEVE GIRVIN and SHRUTI PURI. Shruti has advised projects spanning my

time at Yale from quantum chemistry to large-scale error decoding. I appreciate her commitment to ensuring that experimentalists speak accurately about theory topics. Like Rob, Shruti excels at hiring post-docs and students. Her group's expertise on randomized benchmarking, quantum trajectories, and error correction has been an invaluable resource these past years.

My favorite memory of Steve involves a competition mediated by ISAAC CHUANG, who was a collaborator on my first project at Yale. After a long discussion, Rob, Steve, and Ike arrived at an equation that expressed the fidelity of a certain circuit as a function of the fidelity of its constituent Gaussian gates. A few approximations later, we needed to only solve the quadratic equation to find the target Gaussian gate fidelity. Two camps immediately formed: CHRIS WANG and I on our laptops, and Rob and Steve without. Ike fired the race's starting shot and I quickly pulled up WolframAlpha. Rob began scribbling out the square root via long division while Steve calmly bowed his head to his chest. Less than a minute later, we all had answers. Rob produced the correct result on paper to two digits of precision, but Steve reached three correct digits via mental math alone! We were all stunned, but not surprised. As a young second year student, it was encouraging to see that even the titans of our field need to solve the quadratic equation every now and then.

I'm thankful for PETER RAKICH for joining my committee on short notice and providing valuable feedback. I first met my external reader SERGE ROSENBLUM while he was a post-doc in RSL for the first $\sim$ 2 years of my degree. His wit and precision are both top-tier and make any discussion with him as fun as it is productive. The last experiment in this dissertation is heavily inspired by Serge and PHIL REINHOLD's pioneering work implementing fault-tolerant SNAP with a three-level ancilla transmon. I'm also in the unique position where I need to thank my external reader for fabricating the device on which *two* of the experiments in this dissertation were performed. Seven years later, Smeagol (semi-affectionately called Gollum by

Rob when the T_2 crashes) continues producing science. Finally, I want to express my gratitude to Serge for taking the time to provide such detailed feedback during a difficult time for your country and group.

While I unfortunately had fewer physics discussions with MICHEL DEVORET, I found them all educational. What our conversations lacked in physics, they easily made up for in our shared love of cinema. I had the privilege to teach *Cinema and Physics* with Michel and FRANCESCO CASETTI during the remote teaching era. I immensely enjoyed exploring the impact that physics discoveries had both on the technique and themes of cinema. I've had multiple spontaneous, hour-long conversations with Michel in the hallway about films we had recently seen. Michel's recommendations across film and TV are both extensive and reliable.

No acknowledgement section can omit LUIGI FRUNZIO, who is the friendliest and likely most essential member of Becton. I could write a full chapter of stories involving Luigi ranging from him accidentally speaking Italian during a meeting to picking up deliveries from his home during the peak of Covid. Luigi inducted me into the rolls of peer reviewers, he assisted with several large FPGA orders, and we repaired both Plassys together. I'm grateful for the advice he offered at every stage of my career and for his editing assistance while writing this dissertation.

I had the privilege of collaborating with many excellent scientists during my time in RSL. ZLATKO MINEV welcomed me to the floor my first summer at Yale and taught me many microwave basics. BRIAN LESTER was my first post-doc and began the Franck-Condon Factor experiment that I finished with CHRIS WANG. Brian taught me the principles of quantum optics and the potential of bosonic systems. Chris was a superb mentor for the next few years and made sure that we both learned from every measurement we performed. He encouraged me to do the science that excited me, even when it meant changing experiments. We had a number of excellent collaborators from the Chemistry Department as well including VICTOR BATISTA,

PAT VACCARO, and JESSICA FREEZE.

I had the honor of providing experimental support for YAXING ZHANG, a former post-doc with Steve who is the undisputed master of parametric processes. I owe most of my perturbation theory understanding to him. My next project with CONNOR HANN and SAL ELDER further explored the measurement used in my first paper. Discussions with both deepened my understanding of quantum measurement theory and hidden Markov models.

After a brief yet educational stint with BENJAMIN CHAPMAN, SOPHIA XUE, and STIJN DE GRAAF, I began my last two projects (one covered here) with TAKAHIRO TSUNODA, ALEX DETERS, and BILLY KALFUS. I cannot speak highly enough about my last three years with them. Their love of physics and learning re-energized me after losing some enthusiasm during the isolation of Covid. Taka is the rare physicist who is comfortable in both experiment and theory. I have benefited immensely from his fluency with algorithms and randomized benchmarking. Billy is an equally talented physicist, FPGA engineer, and cannoli critic. I have met few people who understand difficult, technical concepts as quickly as Billy. ALEX DETERS, a fellow Midwesterner and espresso lover, brought a refreshing enthusiasm to lab. His ability to quickly pick up FPGA programming and new physics promises a successful career in graduate school. I hope for another opportunity to work with any of you in the future!

There are many others I would like to thank as well, starting with my classmates ALEC EICKBUSCH, SUHAS GANJAM, JAMES TEOH, SAL ELDER, and SPENCER DIAMOND who shared their knowledge in error correction, device design, materials characterization, and much more. JAHAN CLAES and DANNY WEISS provided theory support on randomized benchmarking and optimal control. I'm thankful to other collaborators and colleagues over the years including NAMITHA LIYANAGE, MARGARET PAVLOVICH, VLAD SIVAK, PATRICK

WINKEL, YUE WU, AKSHAY KOOTTANDAVIDA, CASSADY SMITH, and ANIKET MAITI. I want to also thank TERRI EVANGELISTE, FLORIAN CARLE, RACQUEL MILLER, ALEX BOZZI, GISELLE DEVITO, and MARIA RAO for their stewardship of Applied Physics and YQI.

In addition to the friends and colleagues listed above, I want to thank Maks, Ian, Margot, Lauren, Jon, Diego, Noah, Cyrus, Raghav, Mike, and Erik for your years of friendship. Our adventures in and around New Haven form some of my dearest memories. These last six years would have been more difficult without you, and I thank you for your support and encouragement throughout.

Finally, I want to thank my parents Rob and Sandy Curtis for their unwavering belief in me throughout my education. They encouraged my younger brother Ethan and me to always do our best and be satisfied with that. Early on, it was clear that Ethan was at least as good of a student as me, so I've tried to set the bar as high as possible at each stage. Rather obnoxiously, he has met or exceeded me every time, and is currently earning his Ph.D. in chemistry at Stanford. I'm excited to see where your career takes you. I am grateful to my grandmother Charlotte Curtis for attending my defense and driving me around from one computer store to another as a nerdy child.

Chapter 1

Introduction

1.1 Prologue

The history of the development of quantum mechanics is at this point quite well known. The solutions to the ultraviolet catastrophe and photoelectric effect introduced the notion of quantization: that the allowable configurations of many systems often come in countable sets. The development of quantum mechanics accelerated from there, and within a few decades offered accurate models for atomic spectra, perhaps most famously the hydrogen atom. Quantum field theory and the Standard Model followed, whose demonstrated accuracy so far has only been limited by experimental precision.

Einstein, ever the opponent of quantum non-determinism, reconciled the success of quantum mechanics at these predictions by rejecting the wavefunction as anything more than a probability distribution of outcomes over many systems:

> The attempt to conceive the quantum-theoretical description as the complete description of the individual systems leads to unnatural theoretical interpretations, which become immediately unnecessary if one accepts the interpretation that the description refers to ensembles of systems and not

to individual systems.

—Albert Einstein [Schilpp, 1991]

Measurements of atomic spectra and scattering cross-sections performed at the time mostly probed the ability of quantum mechanics to predict the coupling between different states of an atom and excitations of a field. However, quantum mechanics predicts a much stranger world of superposition and entanglement than can be observed via the ensemble measurement of matrix elements alone. The exploration of this side of quantum mechanics had to wait until the 1990's when a series of technical breakthroughs created platforms that allow the measurement and manipulation of individual quantum systems.

In the meantime, physicists such as Paul Benioff and Richard Feynman proposed in 1980 that one could utilize entanglement among these individual quantum systems to perform computations [Benioff, 1980], and that there may even exist problems at which this "quantum computer" can always outperform a classical one, such as the simulation of other quantum systems in physics and chemistry [Feynman, 1982]. These proposals ignited interest in "quantum computing," and over the next two decades, physicists and computer scientists discovered algorithms for communication [Ekert, 1991, Bennett and Brassard, 2014, Deutsch and Penrose, 1997] and computation [Shor, 1994, Grover, 1996]. The crown jewel of this collection is Shor's algorithm [Shor, 1994], which provides an astonishing exponential speedup in factoring numbers over known classical methods. This discovery established a long-term target for the nascent field of quantum computing and drew the interest of the national security apparatus, whose funding has fueled the field for decades, including this dissertation.

At this point, theoretical advances had far outstripped experimental ones. Indeed, even 30 years later, Shor's algorithm still remains decades away from practical realization. The missing ingredient was the ability to prepare and measure single quantum systems. In the 1950's, Paul invented the radio frequency ion trap [Paul, 1990, Paul

and Steinwedel, 1953] which uses oscillating electric fields to suspend single atoms in a vacuum chamber for individual measurement. Thermal vibrations of the ions in the trap limited the quality of measurements until laser cooling enabled the preparation of ions in their vibrational ground states [Wineland et al., 1978]. The laser trapping and cooling of neutral atoms followed approximately a decade later [Migdall et al., 1985, Lett et al., 1988].

Now that physicists could prepare and measure single quantum systems, the rate of progress increased and the first quantum gate was performed between two qubits [Monroe et al., 1995]. In addition to gates, physicists discovered a novel type of measurement that minimizes the backaction on the quantum system being measured [Braginsky et al., 1980, Brune et al., 1992, Brune et al., 1996]. These accomplishments demonstrated that the fundamental ingredients of a quantum computer are feasible: state preparation, quantum gates, and state measurement. However, they also showed that the rate at which quantum computers experience information-corrupting errors precludes the immediate implementation of the exciting algorithms developed by Shor and Grover. To execute these algorithms, an additional ingredient is needed: quantum error correction. These successes and the challenges ahead began what some have called the "second quantum revolution."

1.2 The second quantum revolution

In 1995, Intel released the Pentium I P54CS with 3.3 million transistors each with a feature size of 350 nanometers. Gordon Moore, one of Intel's founders, claimed in 1975 that the number of transistors (and thus the computational power) in microprocessors would double every two years. Remarkably, the semiconductor industry has upheld this prediction for nearly 50 years now, despite repeated predictions that advancements would slow. The raw computational power of 50 years of exponential

growth is the driving force behind recent advancements in artificial intelligence, such as GPT-4. It is not known exactly how many NVidia A100 GPUs were used to train GPT-4, but many suspect approximately 20,000 were used. At 54 billion transistors per A100, approximately one *quadrillion* (10^{15}) transistors fabricated with a 4 nanometer process participated in its training.

The power of the transistor can likely take us only so far though. There are no known efficient classical algorithms for factoring numbers into their prime components. Quantum chemistry simulations essential for material and drug design remain difficult even with sophisticated approximations that reduce complexity. New generations of large language models require training even larger neural networks than GPT-4. Quantum computers provide a proven exponential speedup at factoring with Shor's algorithm [Shor, 1994], can access larger Hilbert spaces for quantum chemistry, and may even offer speedups for machine learning [Biamonte et al., 2017].

So why hasn't a quantum computer been built yet? The short answer is that the error rates of quantum bits, qubits, far exceed that of classical bits, transistors. There are also difficulties in scaling the number of any type of qubit from tens to hundreds to millions. Qubits are very sensitive quantum systems susceptible to tiny fluctuations in their surrounding environment. Minimizing the sources of these fluctuations and their effects on qubits is a primary research goal in the field.

Since the beginning of the second quantum revolution, new types of quantum bits have emerged as candidates for a quantum computer. One very promising platform is based on the superconducting transmon qubit [Koch et al., 2007], which developed from earlier work on Josephson junction-based qubits [Devoret et al., 1984, Devoret et al., 1985, Martinis et al., 1985, Esteve et al., 1986, Martinis et al., 1987, Nakamura et al., 1999]. Transmons are often referred to as "artificial atoms," since they attempt to replicate the behavior of the atoms (and ions) used in the first experiments on quantum systems, but with the convenience of more accessible energy scales and

tunable properties at the time of manufacture. Since then, transmon lifetimes and on-chip densities have increased to hundreds of microseconds [Ganjam et al., 2023] and 127 transmons per chip [Kim et al., 2023] (although not necessarily both at the same time). Here at Yale, we have successfully used transmons coupled to long-lived superconducting cavities [Reagor et al., 2016] that form the basis of our platforms for error correction.

To overcome the high error rate of qubits, Peter Shor recognized that many physical qubits would need to be combined into a single "logical" qubit [Shor, 1995]. This started the field of quantum error correction, which has proposed many different approaches to achieve a high fidelity qubit ranging from utilizing continuous variable hardware [Gottesman et al., 2001] to the popular surface code [Kitaev, 1997] that tiles many physical qubits into a single logical one. At Yale, our goal is to develop high-fidelity qubits utilizing bosonic modes that contain more than two levels, allowing us to add redundancy at the hardware level before arraying many of these qubits into a surface code, for example.

There are two approaches to quantum error correction (QEC): "feedback" approaches that require measurements of the quantum system followed by actions determined by those measurement results, and "autonomous" schemes, where one engineers the quantum system in such a way that it naturally tends to counter-act the effect of errors. The transmon-cavity architecture provides a unique amount of flexibility in performing the measurements required for feedback error correction and engineering the interactions necessary for "autonomous" error correction. Examples of implementations of the feedback scheme include cat code error correction [Ofek et al., 2016], GKP stabilization [Campagne-Ibarcq et al., 2020, Sivak et al., 2023], binomial code error correction [Ni et al., 2023] while autonomous implementations include cat codes [Gertler et al., 2021, Leghtas et al., 2015, Leghtas et al., 2013, Holland et al., 2015] and Kerr-cat variants [Puri et al., 2020, Puri et al., 2017, Grimm

et al., 2020, Frattini et al., 2022, Venkatraman et al., 2023]. These same measurements and interactions have applications in quantum simulation as well, with exciting applications to chemistry [Wang et al., 2020, Wang et al., 2023].

Arraying large numbers of the qubits mentioned above into a surface code (itself a feedback-based error-correcting code) to make a high-quality logical qubit is a major and ongoing challenge [Acharya et al., 2023]. The main task in the surface code is the measurement of quantities known as stabilizers. These stabilizers determine if an error has occurred in a small neighborhood of qubits, without revealing and spoiling the quantum information encoded in the non-local correlations. To minimize the resources needed to construct one of these logical qubits, the physical qubits and measurements of them must be of as high quality as possible. Additionally, recent proposals have shown that the ability to detect errors during qubit gates and stabilizer measurements significantly reduces hardware requirements [Teoh et al., 2023, Wu et al., 2022]. Whether error detection is used or not, reliable measurements of the quantities that indicate errors is one of the most essential ingredients to feedback-based QEC. Demonstrating the implementation, and advantage of error-detected gates on logical qubits is the primary aim of this dissertation.

1.3 Dissertation structure

Each of the advancements discussed above in the study of quantum systems (and QEC) is a result of breakthroughs in quantum control and measurement. This dissertation will present some new techniques that add to both our control and measurement toolboxes. The chapters are ordered with respect to the complexity and novelty of the measurement technique utilized, building up to a demonstration of an error-detected gate.

As we are focused on quantum control and measurement, we begin in Chapter 2

with a brief review of the current state-of-the-art, including concepts on which we seek to improve. The main building block for bosonic error correction and simulation at Yale is the transmon-cavity pairing. The cavity excels at maintaining the coherence of quantum states, but does not possess enough inherent non-linearity needed for control. On the other hand, the transmon is highly non-linear, rendering it easy to control, but its coherence times are much shorter. When we couple these devices together, the resulting hybridization enables the control of both modes. This dissertation will not provide an in-depth introduction to the physics of the transmon or cQED; excellent synopses can be found in other Yale theses [Schuster, 2007, Chou, 2018, Gao, 2018]. Our review focuses on the Hamiltonian model that describes the composite system, and how we use it to design new controls and measurements. In particular, we review optimal control [Reinhold, 2019, Heeres et al., 2017] and a variety of ways to compute the fidelity of a quantum gate.

The simplest measurement employed in cQED is one that occurs at the end of an experiment. Generally, these end-of-line measurements extract a single bit of information from our transmon qubit (remember that qubit refers to "quantum bit"). In chapter 3, we use end-of-line measurements to characterize the cavity's residual Kerr non-linearity by constructing its Wigner function via single bit measurements. While inherently small, cavity Kerr distorts quantum states and can even produce uncorrectable errors in logical qubits [Campagne-Ibarcq et al., 2020]. We discuss a new technique derived in [Zhang et al., 2022] that tunes Kerr with a simple drive on the transmon. We show that this drive does not spoil system coherences and successfully cancels (or amplifies) Kerr.

There are cases where one may want to measure a quantity that has more than two outcomes. This is especially true in quantum simulation problems, where we use our bosonic cavity to simulate other bosonic modes. Measurements of infinite-dimensional bosonic modes can in principal have infinitely many outcomes. In our

previous work simulating Franck-Condon factors [Wang et al., 2020], we needed to measure in a single shot the number of photons in a cavity. Chapter 4 describes a measurement that extracts multiple bits of information by concatenating four binary QND measurements [Curtis et al., 2021] that measure the bitwise representation of the photon number. We show that the error syndromes of this measurement are predictable and small. Once we calibrate a hidden Markov model describing these errors, we can invert their effect on ensemble measurements. Our error mitigation technique reduces the average error by an order of magnitude.

A large fraction of the infidelity in the bitwise measurement comes from transmon decoherence. The error mitigation technique in Chapter 4 reduces errors in ensembles of measurements, but it cannot detect errors as they occur in real-time. Until recently, we focused on constructing fault-tolerant gates on logical qubits [Rosenblum et al., 2018, Reinhold et al., 2020], which means that a failure in the quantum hardware does not erase any stored quantum information. But in the last year, it was shown that merely detecting the error, a much easier task, provides a substantial benefit for qubits used in surface codes [Wu et al., 2022, Teoh et al., 2023, Kubica et al., 2022]. In Chapter 5 we define an error-detected gate and derive the factor of improvement that error detection can provide. This improvement is limited by the fidelity and QNDness of our measurements, provided that we can construct our gates to fail in a detectable way. In Chapter 6, we discuss what is needed to detect transmon decay and dephasing errors, and introduce the notion of path independence that motivates a new approach to optimal control which optimizes gate fidelity conditional on a transmon measurement after the gate. We optimize this cost function in the presence of up to one dephasing jump, aiming to achieve error detection against a single dephasing event. We focus on a specific gate, a logical Z rotation, as a test case for error detection.

The next two chapters focus on the physical implementation and characterization

of an error-detected logical Z rotation on the binomial code. In Chapter 7, we expand our transmon qubit to a qutrit. If we form a qubit out of the first and third states of the qutrit, any measurements of the qutrit in the second state indicates that a decay (or heating) error occurred [Rosenblum et al., 2018, Reinhold et al., 2020, Elder et al., 2020]. We describe a way to control the $g - f$ qubit and use leakage randomized benchmarking to measure the fidelity of $g - f$ qubit gates. To demonstrate the improvement from error-detected gates, we measure the average logical Z gate fidelity with character randomized benchmarking in Chapter 8. This measurement shows that error detection reduces the infidelity across many different error rates, up to a factor of 22 producing a peak fidelity of 0.9995.

We conclude the dissertation in chapter 9 and discuss some remaining questions and interesting new directions.

Chapter 2

Background and definitions

In this chapter, we review a number of concepts that appear in each of the following chapters. Our goal here is to set the stage for new techniques that we will introduce later that build upon these basics. The topics covered in this chapter include:

- The cQED Hamiltonian
- Simple interactions produced by driving the transmon
- Derivative removal by adiabatic gate (DRAG) for improved transmon gates
- How to compute the fidelity of a gate, and then how to design that gate with optimal control

2.1 cQED background

The fundamental building block of cQED systems at Yale are transmon qubits coupled to 3D superconducting cavities made from high-purity aluminum. When cooled below the critical temperature T_c of aluminum, the bosonic modes of these cavities support long photon lifetimes. This makes them ideal for storing quantum information. In this section, we review the capabilities relevant to this dissertation that transmon-

cavity systems enable. A more exhaustive introduction to the cQED platform can be found in [Blais et al., 2004, Blais et al., 2021, Koch et al., 2007] and many other Yale dissertations [Gao, 2018, Chou, 2018, Schuster, 2007].

2.1.1 Transmon Hamiltonian

The transmon is an anharmonic oscillator whose first two levels are often used as a qubit. Transmons are built from Josephson junctions in which an insulator is sandwiched between two superconductors. The Hamiltonian of a Josephson junction in parallel with a capacitor is

$$H = 4E_C(\hat{n} - n_g)^2 - E_J \cos\hat{\varphi}, \tag{2.1}$$

where $\hat{n}$ is the number of Cooper pairs and $\hat{\varphi}$ is the phase difference across the superconductors and $E_J \gg E_C$ [Koch et al., 2007]. This Hamiltonian is diagonalized up to $O\left(\hat{\varphi}^2\right)$ in terms of creation and annhilation operators $a^\dagger, a$

$$\hat{\varphi} = \left(\frac{2E_C}{E_J}\right)^{1/4}\left(a + a^\dagger\right) = \varphi_0\left(a + a^\dagger\right) \tag{2.2}$$

$$\hat{n} = \frac{i}{2}\left(\frac{E_J}{2E_C}\right)^{1/4}\left(a^\dagger - a\right) \tag{2.3}$$

leaving terms of $O(\varphi_0^4)$ as a perturbative correction in the small parameter φ_0. This produces the Hamiltonian

$$H/\hbar = \omega_a a^\dagger a - \frac{E_J}{\hbar}\left(\cos(\varphi_0(a + a^\dagger)) - \frac{\varphi_0^2(a + a^\dagger)^2}{2}\right). \tag{2.4}$$

where $\omega_a/\hbar = \sqrt{8E_J E_C} - E_C$ is the resonant frequency of the anharmonic transmon oscillator [Blais et al., 2021]. From now one, we set $\hbar = 1$.

2.1.2 Coupling cavities and transmons

Fabricated on small sapphire chips, these devices can be capacitively coupled to both 2D and 3D resonators. For a cavity mode c and a qubit mode a this adds a linear coupling term

$$H = \omega_a a^\dagger a + \omega_c c^\dagger c - \underbrace{E_J \left(\cos\left(\varphi_0 (a + a^\dagger)\right) - \frac{\varphi_0^2 (a + a^\dagger)^2}{2} \right)}_{\text{Transmon potential}} + \underbrace{g \left(c^\dagger a + c a^\dagger \right)}_{\text{JC coupling}}. \tag{2.5}$$

This linear Jaynes-Cummings (JC) coupling terms "dresses" the qubit and cavity modes, meaning that the eigenstates of Eq. 2.5 contain contributions from both modes. We choose to diagonalize this portion of the Hamiltonian and treat the non-linear transmon potential as a perturbation following the procedure in [Blais et al., 2021]. The order of diagonalization is reversed in [Zhang et al., 2022] which is discussed in Chapter 3. The unitary transformation

$$U_\Lambda = \exp\left(\Lambda \left(c a^\dagger - c^\dagger a \right)\right) \tag{2.6}$$

exactly diagonalizes the coupling and hybridizes the eigenstates of each mode resulting in new operators A, C whose excitations participate in both bare modes. After this Bogoliubov transformation, the old mode operators can be expressed in terms of the new ones via

$$U_\Lambda c U_\Lambda^\dagger = \cos(\Lambda)\, C + \sin(\Lambda)\, A = \xi_A C - \xi_C A \tag{2.7}$$

$$\sim C - \frac{g}{\Delta} A \tag{2.8}$$

$$U_\Lambda a U_\Lambda^\dagger = \cos(\Lambda)\, A - \sin(\Lambda)\, C = \xi_A A + \xi_C C \tag{2.9}$$

$$\sim A + \frac{g}{\Delta} C \tag{2.10}$$

where $\Lambda = \frac{1}{2}\arctan\left(\frac{g}{\Delta}\right)$ is the strength of the mode hybridization in terms of the coupling g and detuning $\Delta = \omega_a - \omega_c$. Likewise, we can express the new mode operators in terms of the old:

$$U_\Lambda^\dagger C U_\Lambda = \cos(\Lambda)\, c - \sin(\Lambda)\, a \tag{2.11}$$

$$U_\Lambda^\dagger A U_\Lambda = \cos(\Lambda)\, a + \sin(\Lambda)\, c. \tag{2.12}$$

Notice that for small Λ, the new dressed modes A, C approach the bare modes a, c.

These bare modes aren't generally experimentally accessible; as soon as we introduce a coupling between any two elements, the mode structures deform into the dressed basis as electric fields leak into adjacent mode volumes. This means that we can't design a transmon and cavity in isolation and then bolt them together with predictable outcomes. When both components are designed together, we use tools such as black box quantization [Nigg et al., 2012] and energy participation ratios [Minev et al., 2021] to predict the Hamiltonian parameters based on the electric field modes of the composite package.

Before changing bases, there are two approximations that we make. First, we expand the non-linear transmon potential in $\varphi \ll 1$. We generally keep terms

$$E_J\left(\cos(\varphi_0(a+a^\dagger) - \frac{\varphi_0^2(a+a^\dagger)^2}{2}\right) \sim -\frac{E_J\varphi_0^4}{24}(a+a^\dagger)^4 + \frac{E_J\varphi_0^6}{720}(a+a^\dagger)^6 \tag{2.13}$$

to $O(\varphi_0^4)$, but sometimes higher order terms $O(\varphi_0^6)$ need to be included. For the moment, we neglect the sixth order terms, which are handled in the same way as the fourth order terms. Second, when expanding the $O(\varphi_0^4)$, we normal-order and perform a rotating-wave approximation (RWA) that neglects terms that don't conserve

excitation number

$$-\frac{E_J\varphi_0^4}{24}(a+a^\dagger)^4 \sim -\frac{E_C}{2}a^\dagger a^\dagger aa. \tag{2.14}$$

This approximation produces a Hamiltonian

$$H = \omega_a a^\dagger a + \omega_c c^\dagger c - \frac{E_C}{2}a^\dagger a^\dagger aa + g\left(c^\dagger a + ca^\dagger\right). \tag{2.15}$$

that still captures all of the relevant dynamics in this dissertation, although there may be scenarios where one needs to perform the RWA at a later stage.

Let's now switch bases in Eq. 2.15 from the bare mode operators a, c to the dressed operators A, C and retain only

$$\tilde{H} = U_\Lambda H U_\Lambda^\dagger = \omega_A A^\dagger A + \omega_C C^\dagger C - \frac{E_C}{2}(\xi_A A^\dagger + \xi_C C^\dagger)^2(\xi_A A + \xi_C C)^2 \tag{2.16}$$

with new mode frequencies [Blais et al., 2021]

$$\omega_A = \frac{1}{2}\left(\omega_a + \omega_c + \sqrt{\Delta^2 + 4g^2}\right) \tag{2.17}$$

$$\omega_C = \frac{1}{2}\left(\omega_a + \omega_c - \sqrt{\Delta^2 + 4g^2}\right). \tag{2.18}$$

Expanding the powers in Eq. 2.16 produces every combination $wxyz$ where $w, x \in \{\xi_A A^\dagger, \xi_C C^\dagger\}$ and $y, z \in \{\xi_A A, \xi_C C\}$. In the absence of drives, we continue to neglect the counter-rotating terms that don't conserve energy (equivalent to terms that don't conserve the excitation number in each mode). This produces the standard cQED Hamiltonian

$$H_{\text{cQED}} = \omega_A A^\dagger A + \omega_C C^\dagger C - \underbrace{\frac{\alpha}{2}A^\dagger A^\dagger AA}_{\text{Transmon anharmonicity}} - \underbrace{\frac{K_C}{2}C^\dagger C^\dagger CC}_{\text{Cavity Kerr}} - \underbrace{\chi A^\dagger A C^\dagger C}_{\text{Dispersive shift}} \tag{2.19}$$

where $K_A \sim E_C \xi_C^4$, $\alpha \sim E_C \xi_A^4$, and $\chi \sim \sqrt{2\alpha K_C}$ [Chou, 2018]. The $\sim$ in these expressions emphasizes that these expressions are to leading order in φ_0. Fully diagonalizing the cosine potential incorporates all orders of φ_0 and rotating-wave approximation effects [Koch et al., 2007, Sank et al., 2016].

One higher order effect that commonly appears in experiments is that the dispersive shift is generally not linear in $A^\dagger A$. Here, we rewrite the dispersive shift as

$$\sum_k k\chi_k |k\rangle\langle k| C^\dagger C \tag{2.20}$$

since often the measured value of χ_f differs substantially from χ_e (see Table 7.1 for an example).

2.2 Designing control pulses for gates

Of course, we are interested in performing gates on the transmon and cavity, not just observing the idle evolution of H_{cQED}. We stimulate new interactions by again capacitively coupling the qubit and cavity to microwave cables that connect to arbitrary waveform generators. This coupling $g_{TA}\left(T^\dagger A + T A^\dagger\right)$ and $g_{TC}\left(T^\dagger C + T C^\dagger\right)$ takes the same form as the Jaynes-Cummings coupling, but we can model the transmission line mode T as a large bath of photons that isn't impacted by exchange of a single photon and demote T to a complex number describing the phase and strength of the bath. In this "stiff pump" approximation [Kamal et al., 2009], the microwave control pulses couple to the qubit and cavity via

$$\epsilon_1^*(t) A + \epsilon_1(t) A^\dagger + \epsilon_2^*(t) C + \epsilon_2(t) C^\dagger \tag{2.21}$$

where $\epsilon_1(t), \epsilon_2(t)$ are microwave signals produced by an arbitrary waveform generator.

We implement specific interactions by choosing the frequency and envelope of $\epsilon_1(t)$. There are two main approaches to designing control pulses: time-independent design and optimal control. In time-independent design, we manually choose pulse frequencies and perform rotating frame transformations to produce a time-independent Hamiltonian generator of a specific unitary gate, such as squeezing [Wang et al., 2020, Grimm et al., 2020], beamsplitters [Wang et al., 2020, Gao et al., 2018, Gao et al., 2019, Chapman et al., 2023, Lu et al., 2023], single-photon addition [Rosenblum et al., 2018], and chi-matching [Rosenblum et al., 2018, Reinhold et al., 2020]. The other approach, optimal control, uses the Hamiltonian model to compute waveforms $\epsilon(t)$ that implement any desired unitary [Heeres et al., 2017, Reinhold, 2019].

2.2.1 Time-independent design

Cavity operations

In the presence of microwave drives (often called pumps, in a nod to quantum optics), the Hamiltonian in Eq. 2.16 can implement a variety of operations on the cavity that are otherwise negligibly weak. Notice that we do not start deriving the driven Hamiltonian from H_{cQED} in Eq. 2.19 because a number of RWA's have been made that are no longer valid with drives. Under up to two pumps on the transmon mode A, the driven Hamiltonian is

$$H_d = \tilde{H} + \epsilon_1^*(t)A + \epsilon_1(t)A^\dagger + \epsilon_2^*(t)A + \epsilon_2(t)A^\dagger \tag{2.22}$$

where $\epsilon_1(t), \epsilon_2(t)$ are control functions. Due to the presence of the transmon nonlinearity $\frac{\alpha}{2}A^\dagger A^\dagger AA$, $A + A^\dagger$ does not result in a displacement, as it does in a linear mode. To find the effect of these pumps, we diagonalize the time-dependent $\epsilon_1(t)$

pump terms (without loss of generality) with the unitary transformation

$$U(t) = \exp\left(-\varepsilon_1(t)A^\dagger + \varepsilon_1^*(t)A\right). \tag{2.23}$$

The transformation

$$H_d' = U(t)H_dU^\dagger(t) + i\dot{U}(t)U^\dagger(t) \tag{2.24}$$

implements the basis change $A \to A + \varepsilon_1(t)$ where $\varepsilon_1(t)$ has frequency space solution

$$\varepsilon_1[\omega] = \frac{-i\epsilon_1[\omega]}{\kappa/2 + i(\omega + \omega_A)} \tag{2.25}$$

where $\kappa = 1/T_2$ is the linewidth of the transmon. For a monochromatic drive $\epsilon_1(t) = e^{-i\omega_1 t}\epsilon_1$, the time domain solution is

$$\varepsilon_1(t) = \frac{-i\epsilon_1 e^{-i\omega_1 t}}{\kappa/2 + i(\omega_A - \omega_1)} = \varepsilon_1 e^{-i\omega_1 t}. \tag{2.26}$$

We often prefer to work with monochromatic drives when deriving the frequency conditions for pumped operations as it makes the frequency content explicit. First, make the substitution $A \to A + \varepsilon_1 e^{-i\omega_1 t} + \varepsilon_2 e^{-i\omega_2 t}$ in Eq. 2.22 to enter the displaced frame H_d'

$$H_d' = \omega_A A^\dagger A + \omega_C C^\dagger C - \frac{E_C}{2} \sum_{\substack{w,x\in S^\dagger \\ y,z\in S}} wxyz \tag{2.27}$$

where $S = \{\xi_A A, \xi_C C, \xi_A e^{-i\omega_1 t}\varepsilon_1, \xi_A e^{-i\omega_2 t}\varepsilon_2\}$ and $S^\dagger$ is the element-wise Hermitian conjugate of S. Cavity Kerr $-\frac{K_C}{2}C^\dagger C^\dagger CC$ is still a component of H_d', but the drives provide photons (energy) to overcome previously prohibited interactions.

The choice of drive frequency ω_1, ω_2 determines the terms in H_d' that survive

the rotating wave approximation while expanding Eq. 2.27. We can implement the Gaussian squeezing and beamsplitter operations by setting the frequencies of the two drives to make those interactions resonant. For squeezing, drive frequencies satisfying $2\omega_C = \omega_1 + \omega_2$ bring the squeezing interaction

$$-\frac{2\alpha\xi_C^2\varepsilon_1\varepsilon_2}{\xi_A^2}e^{-i(\omega_1+\omega_2)t}C^2 + \text{h.c.} \tag{2.28}$$

into resonance along with some Stark shifts to produce

$$H_d' = H_{\text{cQED}} - 2\alpha(|\varepsilon_1|^2 + |\varepsilon_2|^2)A^\dagger A - 2\alpha\frac{\xi_C^2}{\xi_A^2}(|\varepsilon_1|^2 + |\varepsilon_2|^2)C^\dagger C - \frac{2\alpha\xi_C^2\varepsilon_1\varepsilon_2}{\xi_A^2}e^{-i(\omega_1+\omega_2)t}C^2 + \text{h.c.}. \tag{2.29}$$

When A is coupled to two cavities C, D, both modes participate in the transmon and Eq. 2.7 generalizes to

$$U_\Lambda a U_\Lambda^\dagger = \xi_A A + \xi_C C + \xi_D D, \tag{2.30}$$

resulting in a driven Hamiltonian

$$H_d' = \omega_A A^\dagger A + \omega_C C^\dagger C + \omega_D D^\dagger D - \frac{\alpha}{2}\sum_{\substack{w,x\in S'^\dagger \\ y,z\in S'}} wxyz \tag{2.31}$$

where $S' = S \cup \{\xi_D D\}$. When the drives are tuned to the (assumed to be large) difference between the cavities, $\omega_1 - \omega_2 = \omega_C - \omega_D$, we bring the beamsplitter interaction into resonance

$$\begin{aligned} H_d' =& H_{\text{cQED}} - 2\alpha(|\varepsilon_1|^2 + |\varepsilon_2|^2)A^\dagger A - 2\alpha\frac{\xi_C^2}{\xi_A^2}(|\varepsilon_1|^2 + |\varepsilon_2|^2)C^\dagger C - -2\alpha\frac{\xi_D^2}{\xi_A^2}(|\varepsilon_1|^2 + |\varepsilon_2|^2)D^\dagger D \\ &- \frac{2\alpha\xi_C\xi_D^*\varepsilon_1^*\varepsilon_2}{\xi_A^2}e^{i(\omega_C-\omega_D)t}CD^\dagger + \text{h.c.}, \end{aligned} \tag{2.32}$$

along with similar Stark shifts as in Eq. 2.29.

In general, pumping the non-linearity of the Josephson junction is a powerful resource that enables a wide range of capabilities including Gaussian operations discussed above [Wang et al., 2020, Gao et al., 2019, Gao et al., 2018], teleported gates [Chou et al., 2018], and communication and entanglement between modes [Burkhart, 2020, Narla et al., 2016, Pfaff et al., 2017, Kurpiers et al., 2019].

Qubit operations

In the previous section, we used the transmon as a device to engineer unitary generators on the cavity mode. But the transmon itself also has a Hilbert space we would like to control. In fact, many of the major industry efforts use the transmon as their primary physical qubit. Our section on optimal control will utilize drives on both the transmon and the cavity simultaneously, so it's valuable to review the driven transmon.

In the absence of coupling to a cavity, but in the presence of a drive, Eq. 2.19 becomes

$$H = \omega_A A^\dagger A - \frac{\alpha}{2} A^\dagger A^\dagger A A + \epsilon^*(t) e^{i\omega_A t} A + \epsilon(t) e^{-i\omega_A t} A^\dagger \tag{2.33}$$

where we have explicitly separated the carrier oscillating at ω_A from the pulse envelope $\epsilon(t)$. Let's first eliminate the carrier frequency by performing a unitary transformation with $U(t) = \exp\left(i\omega_d t A^\dagger A\right)$ and enter a frame rotating with the transmon's energy levels. This produces

$$H' = U(t) H U^\dagger(t) - \omega_d A^\dagger A = \underbrace{(\omega_A - \omega_d)}_{\Delta} A^\dagger A - \frac{\alpha}{2} A^\dagger A^\dagger A A + \epsilon^*(t) A + \epsilon(t) A^\dagger \tag{2.34}$$

where the only explicit time dependence is in the envelope. At this point, it's convenient to truncate the transmon's Hilbert space to three levels and proceed with

the Gell-Mann matrices, which are a set of operators on a three level system that generalize the commutation relations of Pauli matrices. The qutrit Hamiltonian is

$$H = \Delta|e\rangle\langle e| + (2\Delta - \alpha)|f\rangle\langle f| + \epsilon^*(t)\left(|g\rangle\langle e| + \sqrt{2}|e\rangle\langle f|\right) + \text{h.c.} \tag{2.35}$$

$$= \Delta|e\rangle\langle e| + (2\Delta - \alpha)|f\rangle\langle f| + I(t)(\sigma_{ge}^x + \sqrt{2}\sigma_{ef}^x) + Q(t)(\sigma_{ge}^y + \sqrt{2}\sigma_{ef}^y) \tag{2.36}$$

where $\epsilon(t) = I(t) + iQ(t)$. When $\Delta = 0$ and $\alpha \to \infty$, $\sigma_{ge}^x, \sigma_{ge}^y$ are the strongest terms in H, and we can simplify H to a simple driven qubit.

$$H = I(t)\sigma_{ge}^x + Q(t)\sigma_{ge}^y. \tag{2.37}$$

A drive with $Q(t) = 0$ generates rotations about the x axis

$$R_X(\theta) = \exp\left(-i\sigma_x \int I_\theta(t)dt\right) = \exp\left(-i\sigma_x\theta/2\right) \tag{2.38}$$

and likewise $I(t) = 0$ generates rotations about y.

However, there are residual effects from the presence of the third level $|f\rangle$ with finite α. To see this effect, again consider an x rotation where $Q(t) = 0$ and perform a unitary transformation

$$U = \exp(iI(t)(\sigma_{ge}^y + \sqrt{2}\sigma_{ef}^y)/\alpha) \tag{2.39}$$

following the DRAG technique [Motzoi et al., 2009] to find an effective Hamiltonian to first order in $I(t)/\alpha$

$$\begin{aligned} H_{\text{DRAG}} =& I(t)\sigma_{ge}^x - \frac{I(t)^2}{\sqrt{2}\alpha}\sigma_{gf}^x + \left(2\Delta - \alpha - \frac{4I(t)^2}{\alpha}\right)|f\rangle\langle f| \\ &+ \left(\Delta - \frac{2I(t)^2}{\alpha}\right)|e\rangle\langle e| + \left(Q - \frac{\dot{I}(t)}{\alpha}\right)(\sigma_{ge}^y + \sqrt{2}\sigma_{ef}^y). \end{aligned} \tag{2.40}$$

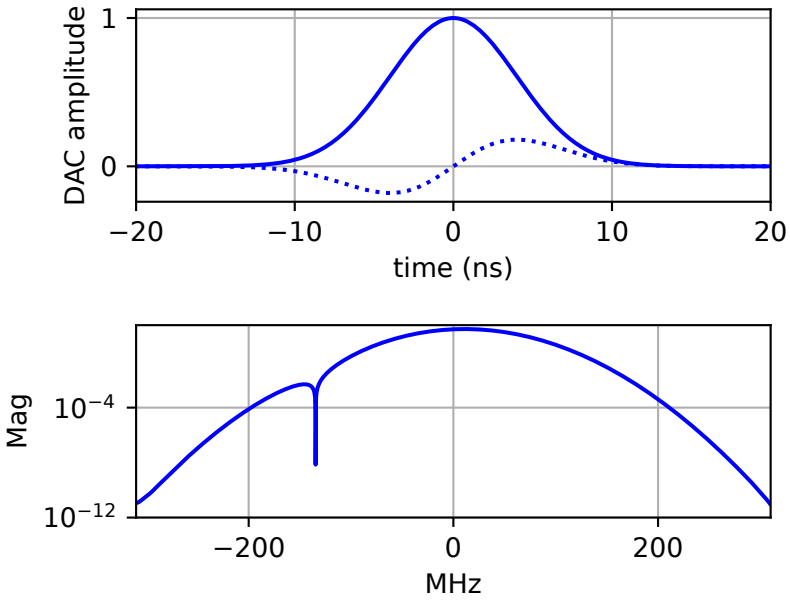


Figure 2.1: Qubit control waveforms with DRAG correction. Our default control pulse $I(t)$ is a Gaussian of the form Eq. 2.45 with $\sigma = 4$ ns and unit DAC amplitude. The waveform of $I(t)$ is the solid line in the **top panel**. We compute $Q(t) = \dot{I}(t)/\alpha \approx 1.19\dot{I}(t)$ for $\frac{\alpha}{2\pi} = 134$ MHz and plot it with dotted lines in the **top panel**. The **bottom panel** is the magnitude of the Fourier transform of the complex waveform $I(t) + iQ(t)$. Our choice of α is motivated by the actual transmon qubit used in Chapters 4, 7, and 8. We do not include the $\Delta(t)$ chirp correction here because in practice a static detuning is usually calibrated. Note the notch that appears in the spectrum at -134 MHz, suppressing the σ^y_{ef} coupling as desired.

The three terms in H_{DRAG} correspond to the rotation we want to implement, a two photon gf coupling off-resonant by α, and an unwanted Stark shift and ef coupling. We can completely cancel the unwanted term by changing our detuning (by chirping the control pulse) $\Delta(t) = 2I^2(t)/\alpha$ and choosing $Q(t) = \dot{I}(t)/\alpha$. See Fig. 2.1 for an example DRAG-corrected qubit control pulse. This technique and higher order corrections have been shown to increase the fidelity of qubit gates on the g, e states [Motzoi and Wilhelm, 2013, Motzoi et al., 2009, Gambetta et al., 2011]. In Chapter 7, we utilize Eq. 2.40 to perform qubit gates on the g, f states.

It's also worth reviewing a simple example of driving the qubit in the presence of

the cavity. Neglecting the cavity Kerr in Eq. 2.19 produces a Hamiltonian

$$H = \omega_A A^\dagger A - \frac{\alpha}{2} A^\dagger A^\dagger A A - \chi A^\dagger A C^\dagger C + \epsilon^*(t) e^{i\omega_A t} A + \epsilon(t) e^{-i\omega_A t} A^\dagger \tag{2.41}$$

with an energy splitting that enables qubit gates conditioned on cavity photon number [Schuster et al., 2007, Krastanov et al., 2015, Reinhold et al., 2020]. Let's quickly see how this arises. First, repeat the rotating frame change that produced Eq. 2.34 with $\omega_d = \omega_A$ and truncate to the first two levels (for simplicity) to arrive at

$$H' = -\chi |e\rangle\langle e| C^\dagger C + \epsilon^*(t) |g\rangle\langle e| + \epsilon(t) |e\rangle\langle g|. \tag{2.42}$$

Notice how this Hamiltonian is block diagonal in the number of photons in the cavity! Now let's enter a frame rotating with the dispersive interaction via the unitary transformation

$$U(t) = \exp\left(-it\chi |e\rangle\langle e| C^\dagger C\right) \tag{2.43}$$

to get a Hamiltonian in the frame of the dispersive interaction:

$$H_\chi = \epsilon^*(t) \exp\left(it\chi C^\dagger C\right) |g\rangle\langle e| + \epsilon(t) \exp\left(-it\chi C^\dagger C\right) |e\rangle\langle g|. \tag{2.44}$$

Now consider the case when the drive is a Gaussian centered at DC with width σ_t and total length 4σ. Properly normalized to $\int_0^{4\sigma} \epsilon(t) dt = \frac{\pi}{2}$ gives

$$\epsilon(t) = \frac{1}{\operatorname{erf}(1/\sqrt{2})} \sqrt{\frac{1}{8\pi}} e^{-(t-2\sigma)^2/2\sigma^2}. \tag{2.45}$$

With this waveform, the Hamiltonian

$$H_\chi = \sum_n \epsilon(t)\cos(\chi n t)|n\rangle\langle n| \otimes \sigma_x + \epsilon(t)\sin(\chi n t)|n\rangle\langle n| \otimes \sigma_y \tag{2.46}$$

implements what we call a selective π-pulse qubit rotation that rotates the qubit conditioned on there being zero photons in the cavity, the only photon number where the oscillations in H_χ vanish. The selectiveness of the π-pulse is determined by χ and the length of the pulse. In Fig. 2.2, we explore the dependence of selectivity on the pulse duration for $\chi/(2\pi) = 1$ MHz. This procedure can be generalized to rotate the qubit conditioned on any set of Fock states by constructing a frequency comb with components at $-n\chi$. This technique can be used to measure cavity parity, or any parity generalization.

2.2.2 Optimal control

In the preceding section, we manually engineered Hamiltonians H that generate a desired unitary $U = \mathcal{T}e^{-i\int H(t)dt}$ where $\mathcal{T}$ is the time-ordering operator. But there are many other gates for which these manual constructions are not obvious. For example, how does one choose a $\epsilon(t)$ that produces a unitary that creates a photon from the vacuum state $U|g\rangle\otimes|0\rangle = |g\rangle\otimes|1\rangle$? These situations require a new approach: optimal control [Khaneja et al., 2005, de Fouquieres et al., 2011].

Optimal control is a set of techniques for finding system controls that achieve a desired outcome. If a system, whether it be quantum [Heeres et al., 2017] or classical, has an accurate model, we can optimize controls against a cost function that defines a distance between the control effects and the desired outcome. In quantum computing, our controls are microwave pulses generated by AWGs and our response model is the

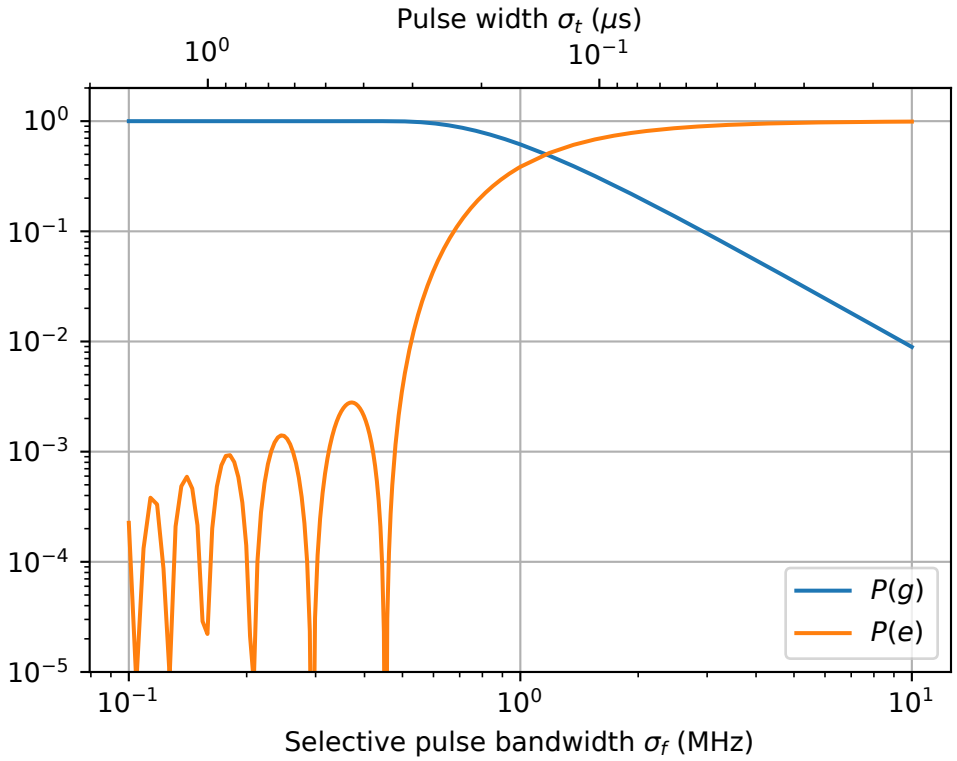


Figure 2.2: Selective π-pulse fidelity. Let's explore the effect of our choice of σ for $\epsilon(t)$ as defined in Eq. 2.45. Consider an experiment where we want to check if the cavity is in vacuum at the end, such as a cavity T_1 measurement. We map the answer to "does the cavity have zero photons?" onto the state of the qubit via H_χ defined in Eq. 2.46. If there are zero photons, the qubit is excited to $|e\rangle$, whereas it remains in $|g\rangle$ if there is one photon in the cavity. The above figure shows the probability of measuring the qubit in g, e when there is one photon in the cavity. Here, $\chi/(2\pi) = 1$ MHz and the pulse is $4\sigma_t$ in total duration. For $\sigma_t > 300$ ns, the probability of an erroneous excitation is low, but as σ_t decreases and the bandwidth $\sigma_f = 1/(2\pi\sigma_t)$ exceeds 600 kHz, the probability of a false positive exceeds 10%. As the pulse bandwidth (duration) further increases (decreases), the qubit pulse becomes unselective, rotating into $|e\rangle$ regardless of the cavity state.

driven cQED Hamiltonian Eq. 2.19

$$H = -\frac{\alpha}{2}A^\dagger A^\dagger AA - \frac{K_C}{2}C^\dagger C^\dagger CC - \chi A^\dagger AC^\dagger C + \epsilon_A^*(t)A + \epsilon_A(t)A^\dagger + \epsilon_C^*(t)C + \epsilon_C(t)C^\dagger \tag{2.47}$$

where we have now entered a frame rotating at ω_A with A and ω_C with C. Note that we have included drives on each mode and entered the co-rotating frame for convenience.

The waveforms $\epsilon_{A/C}(t)$ that reach the qubit and cavity are reconstructed at sampling rate f_s from a finite set of samples $\epsilon_{A/C}^k$ spaced in time by $1/f_s$. We place lowpass filters on the outputs of our AWGs to limit signal bandwidth. This means that as long as bandwidth of $\epsilon_{A/C}(t)$ is below $f_s/2$, the sampled pulse contains as much information as the continuous pulse via the Nyquist-Shannon sampling theorem. This also means that we can choose to optimize the sampled amplitudes rather than some continuous function basis. Our goal is then to find the samples $\epsilon_{A/C}^k$ that produce our desired unitary U. We model the evolution of the gate as a series of "step propagators" U_i

$$|\psi_{\text{initial}}\rangle \; \boxed{U_1} \; \boxed{U_2} \; \cdots \; \boxed{U_N} \; |\psi_{\text{target}}\rangle$$

defined as

$$U_k = e^{-iH_k dt} \tag{2.48}$$

treating the Hamiltonian's time dependence as piecewise-constant

$$H_k = -\frac{\alpha}{2}A^\dagger A^\dagger AA - \frac{K_C}{2}C^\dagger C^\dagger CC - \chi A^\dagger AC^\dagger C + (\epsilon_A^k)^* A + \epsilon_A^k A^\dagger + (\epsilon_C^k)^* C + \epsilon_C^k C^\dagger. \tag{2.49}$$

In practice, we can choose to optimize the amplitudes $\epsilon_{A/C}^k$ or the frequency com-

ponents $\zeta^k_{A/C}$ to add native bandwidth limits beyond $f_s/2$ by setting the Fourier components that exceed our cutoff to zero. $\epsilon^k_{A/C}$ is then the inverse discrete Fourier transform

$$\epsilon^k_{A/C} = \frac{1}{N}\sum_{k=0}^{N-1} \zeta^k_{A/C} \exp\left(\frac{2\pi i}{N} kn\right) \tag{2.50}$$

using the electric engineering convention for the Fourier transform. Other approaches directly optimize the samples $\epsilon^k_{A/C}$ and add bandwidth restrictions via cost functions [Heeres et al., 2017].

Our decision to optimize in frequency space comes with one drawback: $\epsilon^k_{A/C}, \epsilon^{N-1}_{A/C}$, the first and last samples, are not guaranteed to be zero. In fact, it is not possible for a waveform to be fully time and bandwidth limited; one range must extend to infinity. Discrete prolate spheroidal sequences (commonly called Slepians) are a basis of functions that are maximally time and band-limited [Frey et al., 2017, Norris et al., 2018], but these are overkill for our situation. We instead add a Tukey window (also known as a cosine window) with an adjustable risetime to ensure that our pulses start and end at zero amplitude. This broadens the pulse's bandwidth beyond our original window, but only minimally. See Fig. 2.3 for an example.

Now that we have defined the step propagators U_k, we can write the total unitary generated

$$U = U_N \ldots U_2 U_1. \tag{2.51}$$

Our next task is to develop a cost function that evaluates how well our circuit with waveform samples $\epsilon^k_{A/C}$ approximates a target unitary U_{target}.

Unitary cost functions

The most general definition of fidelity between a target unitary U_{target} and the quantum channel $\mathcal{E}$ implementing it is the average gate fidelity [Nielsen, 2002, Bowdrey

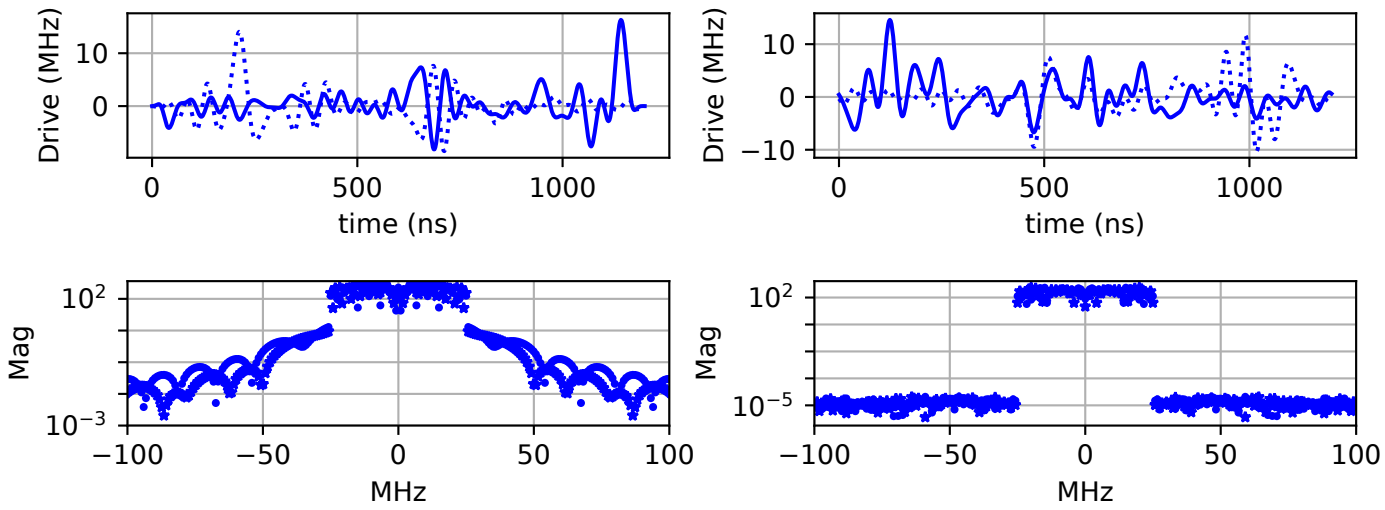


Figure 2.3: Optimal control pulses and frequency content. Shown here are two different pulses optimized to implement a logical Z rotation on a binomial/kitten qubit. **Top row**: waveforms of pulses 1200 ns long restricted to 25 MHz of bandwidth. The pulse on the **left** is multiplied by a Tukey window with ramp 20 ns while the pulse on the **right** has no ramp. Solid circles in the **bottom row** plots are the magnitude of frequency component of the real part of the pulse, while stars correspond to the imaginary part. Since the right pulse is optimized in frequency space, without a ramp there is a brick wall filter at 25 MHz and the frequency content immediately falls by seven orders of magnitude. Unfortunately, this pulse does not start and end at zero amplitude, as shown in the top panel (solid line is real part, dotted imaginary), but the pulse with the Tukey window does.

et al., 2002]

$$\bar{F}_g(U_{\text{target}}, \mathcal{E}) = \int d\psi \, \langle\psi| \, U_{\text{target}}^\dagger \mathcal{E}(|\psi\rangle\langle\psi|) U_{\text{target}} \, |\psi\rangle \, . \tag{2.52}$$

$\bar{F}_g$ is the most complete cost function for optimizing the controls to approximate U_{target}, but it is simpler and often sufficient to consider only the unitary evolution. In the case that $\mathcal{E}$ is unitary, with $\mathcal{E}(\rho) = U\rho U^\dagger$ the average gate fidelity is

$$\bar{F}_g(U_{\text{target}}, \mathcal{E}) = \frac{\left|\text{Tr}\left(U_{\text{target}}^\dagger U\right)\right|^2 + d}{d^2 + d}, \tag{2.53}$$

where d is the dimension of the Hilbert space on which we want to implement U_{target} [Nielsen, 2002, Bowdrey et al., 2002, Emerson et al., 2005]. This means that maximizing $\bar{F}_g$ is equivalent to maximizing the trace that appears in the numerator.

Let's now call the trace in the numerator of Eq. 2.53 the trace fidelity [Schulte-

Herbrüggen et al., 2011, Ball et al., 2021]

$$\mathcal{F}(U_{\text{target}}, U) = \left| \frac{1}{\text{Tr}\,(P)} \text{Tr}\left(P U_{\text{target}}^{\dagger} U\right) \right|^2 . \tag{2.54}$$

This quantity is maximized and unity when $U = U_{\text{target}}$ where P is a projector into some computational subspace $C = \text{Image}(U_{\text{target}})$. We include P since we generally need optimize in a Hilbert space with larger dimension than U_{target}. An advantage of defining U_{target} on a subspace C is that it provides the optimizer with the freedom to choose the effect of U on the non-computational subspace $H \setminus C$ (often random, yet still unitary), often enabling faster convergence and higher fidelities. This allows us to define unitaries on one-dimensional subspaces to prepare Fock states from a qubit in its ground state $U_{\text{target}} |0, g\rangle = |n, g\rangle$ and $P = |0, g\rangle \langle 0, g| + |n, g\rangle |n, g\rangle$ without specifying behavior on the rest of H.

When we rewrite the trace in $\mathcal{F}$ as a sum over a set of orthonormal initial states $\{|\psi_k\rangle\}_k$ that span C with target states $\{|\psi_k^{\text{target}}\rangle = U_{\text{target}} |\psi_k\rangle\}_k$ (note that $|\psi_k^{\text{target}}\rangle$ may carry a phase factor, it is essential to keep this phase, else this is no longer identical to the trace in 2.54), calculating the trace fidelity is reduced to computing N state overlaps

$$\mathcal{F} = \left| \frac{1}{\text{Tr}\,(P)} \sum_k \langle \psi_k^{\text{target}} | \, U \, |\psi_k\rangle \right|^2 . \tag{2.55}$$

Note that the summands in $\mathcal{F}$ are complex numbers whose phases are equal, but not necessarily zero, when $\mathcal{F}$ is maximized. This means that U applies a global phase, but preserves relative phases in a superposition under U. One can instead force the phase of the matrix elements to zero by replacing $|\ldots|^2$ in $\mathcal{F}$ by simply taking the real part instead [Schulte-Herbrüggen et al., 2011]

$$\mathcal{F}^R = \Re\left(\frac{1}{\text{Tr}\,(P)} \text{Tr}\left(P U_{\text{target}}^{\dagger} U\right) \right). \tag{2.56}$$

However, there exist situations where preserving relative phases between superpositions of states is not required. For example, the photon number-resolved measurement we will discuss in Chapter 4 projects the cavity into a Fock state by measuring a series of binary observables whose outcomes do not depend on the superposition phases. In these cases, we can use an "incoherent" cost function that ignores the relative phases imparted by U:

$$\mathcal{F}_j^{\text{incoherent}} = \frac{1}{N} \sum_k \left| \langle \psi_k^{\text{target}} | \, U \, | \psi_k \rangle \right|^2 . \tag{2.57}$$

This provides the optimizer with an additional degree of flexibility that can aid convergence.

Non-unitary cost functions

The case when $\mathcal{E}$ is not unitary is important as well, especially for characterizing gate fidelities in simulation, or optimizing in the presence of decoherence. Thankfully, there is a way to decompose Eq. 2.52 into an average over state transfers as well. The first step replaces the integral over states with a sum over a basis of unitaries B_j

$$\bar{F}_g(U_{\text{target}}, \mathcal{E}) = \frac{\sum_j \text{Tr}\left(U_{\text{target}} B_j^\dagger U_{\text{target}}^\dagger \mathcal{E}(B_j) \right) + d^2}{d^2(d+1)} \tag{2.58}$$

where $\text{Tr}(B_j) = d$ [Nielsen, 2002, Bowdrey et al., 2002]. The minimum value of this fidelity $1/(d+1)$ is nonzero because quantum channels $\mathcal{E}$ are defined to be trace-preserving, which places a lower bound on how much they can scramble a state. Elements of this basis are not pure states, but we can express the B_j in terms of pure

states [Nielsen, 2002]. The following basis states span all B_j

$$\rho_j = |j\rangle\langle j| \tag{2.59}$$

$$\rho_{jk}^{\pm} = \frac{1}{2}(|j\rangle\langle j| + |k\rangle\langle k| \pm |k\rangle\langle j| \pm |j\rangle\langle k|) \tag{2.60}$$

allowing us to rewrite the fidelity of a qubit gate $d = 2$ as an average of these d^2 states

$$\bar{F}_g(U_{\text{target}}, \mathcal{E}) = \frac{1}{4} \sum_{\rho \in \{|+\rangle\langle +|, |-\rangle\langle -|, |0\rangle\langle 0|, |1\rangle\langle 1|\}} \text{Tr}\left(U_{\text{target}} \rho^{\dagger} U_{\text{target}}^{\dagger} \mathcal{E}(\rho)\right). \tag{2.61}$$

We utilize this result from [Nielsen, 2002, Bowdrey et al., 2002] in Chapter 5. The case for $d > 2$ requires expressing each B_j in the basis of $\rho_j, \rho_{jk}^{\pm}$, but we can estimate the fidelity by simply averaging over this basis

$$\bar{F}_g(U_{\text{target}}, \mathcal{E}) = \frac{1}{d^2} \sum_{\rho \in \{\rho_j\} \cup \{\rho_{jk}^{+}\} \cup \{\rho_{jk}^{-}\}} \text{Tr}\left(U_{\text{target}} \rho^{\dagger} U_{\text{target}}^{\dagger} \mathcal{E}(\rho)\right). \tag{2.62}$$

We end by noting that the trace computed above can be re-written as the probability of measuring a pure state $\rho_{\text{target}} = U_{\text{target}} \rho^{\dagger} U_{\text{target}}$ after evolution of ρ under the channel $\mathcal{E}$. We recognize that $\rho_{\text{target}} = |\psi_{\text{target}}\rangle\langle\psi_{\text{target}}|$ which is a projector into a pure state $|\psi_{\text{target}}\rangle$ equal to the measurement effect $E_{|\psi_{\text{target}}\rangle}$. That is, $\bar{F}_g(U_{\text{target}}, \mathcal{E})$ is the average probability

$$\bar{F}_g(U_{\text{target}}, \mathcal{E}) = \frac{1}{d^2} \sum_{\rho \in \{\rho_j\} \cup \{\rho_{jk}^{+}\} \cup \{\rho_{jk}^{-}\}} \text{Tr}\left(E_{|\psi_{\text{target}}\rangle} \mathcal{E}(\rho^{\dagger})\right). \tag{2.63}$$

that the channel $\mathcal{E}$ produces the correct output for any input state.

Optimization

Now that we have a cost function to evaluate the accuracy of a particular choice of samples $\epsilon^k_{A/C}$ (or Fourier components), we optimize our choice via gradient descent. We need to numerically or analytically calculate the gradient $\nabla_{\epsilon^k_{A/C}}\mathcal{F}$ and pass it to an optimizer such as L-BFGS [Liu and Nocedal, 1989] or Adam [Kingma and Ba, 2017]. Previously, this gradient was computed analytically [Reinhold, 2019, Heeres et al., 2017], but since then autodifferentiation (AD) has become popular [Abdelhafez et al., 2019, Eickbusch et al., 2022, Song et al., 2022, Leung et al., 2017]. AD operates by storing the intermediate value of all computations, which is used to compute partial derivatives via the chain rule during back-propagation. This technique is efficient in time, but less efficient in memory. Crucially, it allows us to optimize cost functions whose analytic gradients are hard to compute, which is especially useful if we use Monte-Carlo methods [Abdelhafez et al., 2019] to compute expectation values like $\mathrm{Tr}\left(U_{\mathrm{target}}\rho^\dagger U^\dagger_{\mathrm{target}}\mathcal{E}(\rho^\dagger)\right)$.

Chapter 3

Kerr engineering

End-of-line measurements are probably the most common form of measurement in quantum information. For many platforms, they are the only form of measurement, such as photonics, where single photon detectors (SPDs) convert photons into a classically measurable electric current. At the end, there is no photon left, so the quantum system has been destroyed during the measurement.

In cQED, end-of-line measurements decode the results of quantum circuits, measure photon number occupation, and characterize and calibrate each aspect of our systems. This section discusses using end-of-line measurements to characterize Kerr, the static cavity non-linearity induced by the coupling to the transmon. We discuss several methods to extract K_C and then introduce a method to tune its value via a drive on the transmon.

The static non-linearity of the cavity K_C in Eq. 2.19 induces error in the Gaussian operations discussed in Section 2.2.1 and in any idle time in our experiment. In the binomial code, which has codewords

$$|0_L\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |4\rangle) \tag{3.1}$$

$$|1_L\rangle = |2\rangle \tag{3.2}$$

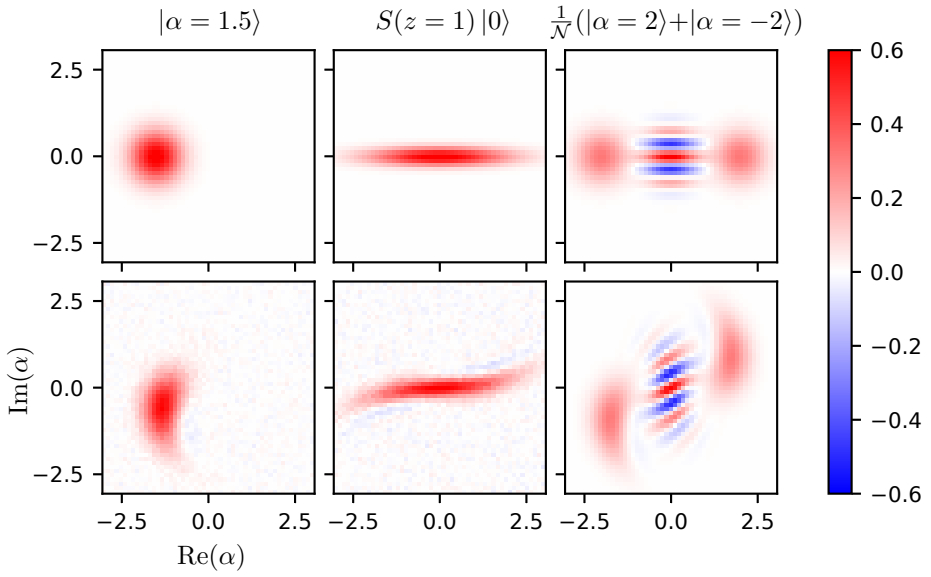


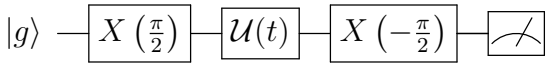
Figure 3.1: **Top row**: Wigner functions $W(\alpha)$ of the effect of $K = 3$ kHz on three different ideal states. **Bottom row**: measured Wigner fucntions. For the coherent state on the **left**, the measured Wigner function shows the effect of 10 μs of idle evolution in the presence of Kerr. In the **middle**, the measured squeezed state experiences Kerr during the squeezing drives $C^2 + \left(C^\dagger\right)^2$ that distorts the state resulting in a "spiral galaxy" effect. The bottom figure in the **right** cat state column is a simulation of 20 μs of idle evolution in the presence of Kerr. No measurements of this state were taken.

Kerr manifests as leakage from $|0_L\rangle \rightarrow \frac{1}{\sqrt{2}}(|0\rangle - |4\rangle)$. It also limits the fidelity of beamsplitters [Wang et al., 2020] and GKP state stabilization [Campagne-Ibarcq et al., 2020, Sivak et al., 2023]. Fig. 3.1 shows some examples of how Kerr distorts idling states or the creation of Gaussian states. For these reasons, it is important to calibrate Kerr, and in some cases employ measures to mitigate its effect.

We also note that Kerr does not have to be the enemy! In fact, one can use Kerr and squeezing together to stabilize cat states [Puri et al., 2017, Grimm et al., 2020, Frattini et al., 2022, Venkatraman et al., 2023]. These efforts are beyond the scope of this dissertation, and we will focus on characterization and mitigation here.

3.1 Characterizing Kerr with Ramsey-like experiments

Ramsey-style measurements are common for cases in which we need to learn about phase coherence and detuning. For a qubit, this circuit is:



where $\mathcal{U}(t)$ is some time evolution that includes decoherence and noise, which affect state superpositions. By sweeping t, we extract quantities such as T_2 and qubit detuning.

We can perform these measurements on cavity states as well, but recall that the linearity of microwave cavity oscillators makes it difficult to address just two states, as in a qubit. Preparing superpositions of Fock states in the cavity for Ramsey experiments generally requires using an optimal control pulse. But, optimizing these pulses requires knowledge of the static Hamiltonian, which includes Kerr. In this section, we discuss a Ramsey-like experiment that utilizes coherent states instead of Fock states to get an estimate of Kerr that can be improved with a subsequent re-run with Fock states.

3.1.1 Coherent states

The microwave cavities used in this dissertation have high intrinsic dephasing times $T_\phi > 10$ ms [Rosenblum et al., 2018] and single-photon lifetimes $T_1 > 1$ ms [Reagor et al., 2016]. Recent works have further optimized the cavity design and achieved single-photon lifetimes $T_1 > 25$ ms with staggeringly high pure dephasing rates $T_\phi > 0.5$ s [Milul et al., 2023]. These long coherences allow us to approximate the evolution $\mathcal{U}(t)$ in a cavity for $t \ll 1$ ms as a unitary $U(t)$ generated by

$$H_{\text{Kerr}} = \delta C^\dagger C - \frac{K_C}{2} C^\dagger C^\dagger C C \tag{3.3}$$

with some possible residual detuning δ. The easiest state to prepare in the cavity with high fidelity is a displaced state

$$D(\alpha) = \exp(\alpha C^\dagger - \alpha^* C) \tag{3.4}$$

$$D(\alpha)\,|0\rangle = e^{-|\alpha|^2/2} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}\,|k\rangle \tag{3.5}$$

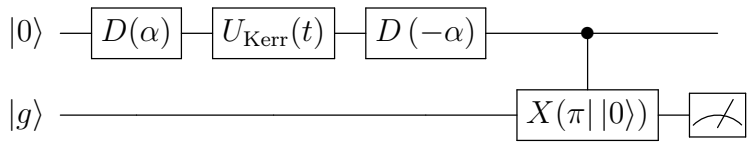
generated simply with drives $\epsilon_C^*(t)C + \epsilon_C(t)C^\dagger$ on the cavity. Displaced states $|\alpha\rangle$ are not eigenstates of H_{Kerr}, but H_{Kerr} does preserve photon number, so we expect the evolution $U(t) = e^{-iH_{\text{Kerr}}t}$ to rotate and smear out the coherent state in some way, especially at short times. Let's check this intuition by finding the overlap of $U(t)\,|\alpha\rangle$ with some other coherent state $|\alpha e^{-i\theta}\rangle$ remembering that $\langle m|n\rangle = \delta_{m,n}$

$$\langle \alpha e^{-i\theta}|U(t)|\alpha\rangle = e^{-|\alpha|^2} \left(\sum_{m=0}^{\infty} \frac{\left(\alpha^* e^{i\theta}\right)^m}{m!}\,\langle m| \right) \left(\sum_{n=0}^{\infty} \frac{\alpha^n}{n!} e^{-i(\delta n - \frac{K_C}{2}(n^2-n))t}\,|n\rangle \right) \tag{3.6}$$

$$= 1 + e^{-|\alpha|^2} \sum_{k=1}^{\infty}\sum_{n=0}^{\infty} \frac{|\alpha|^{2n}}{n!k!} \left(in\left(\theta - \left(\delta + \frac{K_C}{2}\right)t\right) + \frac{iK_C}{2}n^2 t \right)^k \tag{3.7}$$

$$\sim 1 + i|\alpha|^2 \left(\theta - \delta t + \frac{K_C}{2}|\alpha|^2 t\right). \tag{3.8}$$

The fidelity $|\,\langle \alpha e^{-i\theta}|U(t)|\alpha\rangle\,|^2$ is maximized to first order when $\theta(t) = \left(\delta - \frac{K_C}{2}|\alpha|^2\right)t$ for $\frac{K}{2}|\alpha|^2 \ll 1$. This means that $U(t)\,|\alpha\rangle$ largely remains a coherent state, but with a time-dependent phase $|\alpha e^{-i\left(\delta - \frac{K_C}{2}|\alpha|^2\right)t}\rangle$. A displacement of $D\left(-\alpha e^{-i\theta(t)}\right)$ (although we won't know $\theta(t)$ a priori in an experiment) will bring this state back to vacuum $|0\rangle$, which we can check with the selective pulses described in Section 2.2.1 and Fig. 2.2. Our Ramsey Kerr measurement is then



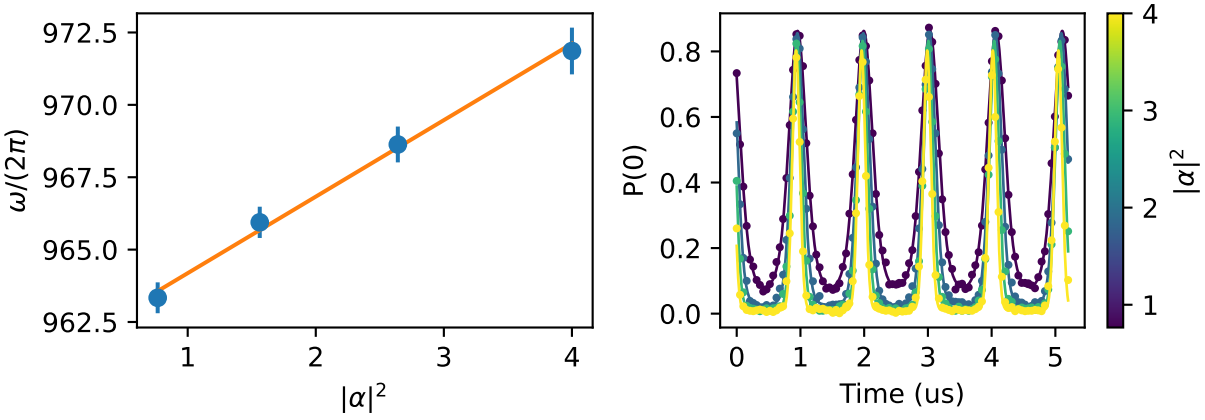


Figure 3.2: To estimate K_C, we first measure $P(|0\rangle)$ for a range of α and $N = 101$ times $t \in [0, t_{\max} = 5.2\mu\text{s}]$ with interval $dt = \frac{\max}{N} = 52\text{ns}$. For each α, we fit $e^{2|\alpha|^2(\cos(\omega t)-1)}$ shown in the **right** plot and extract ω. To find K_C and δ, we fit $\omega = \delta - \frac{K_C}{2}|\alpha|^2$ in the **left** plot as a function of $|\alpha|^2$. This fit yields $K_C/2\pi = 2.64 \pm 0.15$ kHz and $\delta/2\pi = 961.54 \pm 0.32$ kHz. Note that a small difference in oscillation frequency is only apparent after multiple periods. The rate and duration at which we sample allows distinguishing frequencies f between $\left(\frac{1}{2Ndt} = 95\text{kHz}\right) < f < \left(\frac{1}{2dt} = 9.6\text{MHz}\right)$. This motivates our choice of δ, which is implemented via a "digital detuning" that changes the local oscillator frequency of our AWGs.

where $X(\pi|\,|0\rangle)$ indicates a π-pulse selective on zero photons in the cavity. The probability of measuring the cavity in $|0\rangle$

$$P(|0\rangle) = |\,\langle 0|D(-\alpha)U(t)D(\alpha)|0\rangle \tag{3.9}$$

$$\sim |\,\langle \alpha|\alpha e^{-i\left(\delta - \frac{K_C}{2}|\alpha|^2\right)t}\rangle\,|^2 \tag{3.10}$$

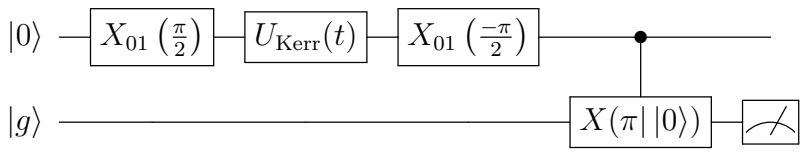
$$= e^{2|\alpha|^2\left(\cos\left(\left(\delta - \frac{K_C}{2}|\alpha|^2\right)t\right)-1\right)} \tag{3.11}$$

oscillates as a function of both t and α. To extract K_C and δ, we scan both t and α and perform two fits as shown in Fig. 3.2.

3.1.2 Fock states

Once we have an estimate of K_C via the coherent state Ramsey-like experiment, we can create OCT pulses to initialize Fock states and refine K_C via a similar Ramsey-

like experiment. We generally do this in two steps, first calibrating δ via a typical cavity T_2 experiment

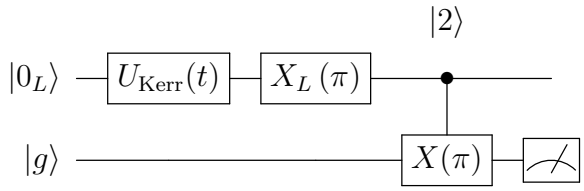


with rotations

$$X_{01}(\theta) = \cos(\theta/2)|0\rangle\langle 0| + \cos(\theta/2)|1\rangle\langle 1| - i\sin(\theta/2)|0\rangle\langle 1| - i\sin(\theta/2)|1\rangle\langle 0| \quad (3.12)$$

defined on the $0, 1$ photon number space. In this subspace, the Kerr term $-\frac{K_C}{2}C^\dagger C^\dagger CC$ vanishes and the frequency we extract in the $P(0)$ oscillations corresponds exactly to δ.

Typically to extract Kerr, we optimize $X_{02}(\theta)$ on the 0-2 photon number subspace and extract K_C from the oscillations in $P(0)$. Other subspaces, such as 0-4 can be used as well. In Fig. 3.3, we use the $|0_L\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |4\rangle)$ logical codeword of the binomial code to better measure K_C. In this case, the measurement circuit is



where the $X_L(\pi)$ maps $|0_L\rangle \to |1_L\rangle = |2\rangle$ which we can easily measure with a selective π-pulse conditioned on two photons in the cavity. Another option is to optimize a unitary that maps $|g\rangle \otimes |0_L\rangle \to |g\rangle \otimes |0_L\rangle$ and $|g\rangle \otimes |1_L\rangle \to |e\rangle \otimes |1_L\rangle$. This approach is not appreciably simpler because it still requires drives on both the transmon and cavity. Furthermore, it's difficult to estimate the fidelity of such a map, whereas we can characterize the logical rotation $X_L(\pi)$ with randomized benchmarking and

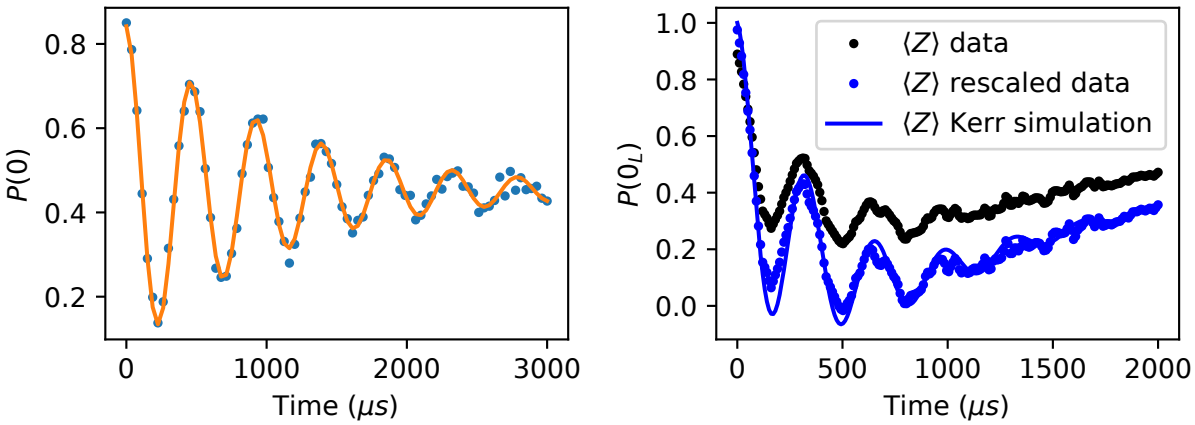


Figure 3.3: **Left**: cavity Ramsey experiment utilizing a superposition of zero and one photon states with fitted $T_2 = 1.11 \pm 0.03$ ms. **Right**: a lifetime measurement of the binomial codeword $|0_L\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |4\rangle)$. Since $|0_L\rangle$ is a superposition of Fock states, it can also be used for Ramsey experiments. There is no added detuning here; the oscillations in $\langle Z_L \rangle$ are a result of Kerr $K_C = 507$ Hz. The envelope decay is due to cavity T_1. To find Kerr, we first rescale the measured expectation value $\langle Z_L \rangle$ to account for logical measurement infidelity. The fitted preparation and measurement contrast is 0.68, limited by the logical $R_Y(\pi)$ rotation used in both state preparation and measurement. We then fit the rescaled data to a master equation simulation which includes cavity photon loss. For states $|\psi\rangle$ outside of the codespace, $\langle \psi | Z_L | \psi \rangle = 0$. This prevents contamination of the measurement by leakage. The master equation simulation captures errors within the codespace and the eventual convergence to zero photons at long delays. The best version of this measurement would shorten the duration to a window where cavity decay is unlikely and add an additional detuning.

the contrast of the selective π-pulse on vacuum. This technique is also amenable to an approach discussed in Section 8.3 that uses a three-level transmon to improve measurement contrast. Finally, note that we can exploit normal-ordering and extend this technique to measure the strength of any non-linearity $(C^\dagger)^n C^n$ by successively preparing and performing Ramsey experiments with superpositions $|0\rangle + |n\rangle$.

3.2 Characterizing Kerr with Wigner functions and density operators

In the previous section we designed interference experiments that measure frequency components corresponding to non-linearities in the cavity. Another, more brute force approach, is to reconstruct the complete quantum state ρ at the end of some evolution

$$\rho = \mathcal{U}(t)\rho_0. \tag{3.13}$$

With access to the final state ρ, we can optimize the fidelity of ρ to some test state

$$\rho(\vec{\lambda}) = \mathcal{U}_{\vec{\lambda}}(t)\rho_0, \tag{3.14}$$

where the evolution $\mathcal{U}_{\vec{\lambda}}(t) = \mathcal{T}e^{\int \mathcal{L}_{\vec{\lambda}}dt}$ is generated by the quantum Liouville superoperator

$$\mathcal{L}_{\vec{\lambda}}[\rho] = i[\rho, H(\vec{\lambda})] + \sum_l \kappa_l \left(J_l \rho J_l^\dagger - \frac{1}{2}\{J_l^\dagger J_l, \rho\} \right) \tag{3.15}$$

for decoherence operators J_l with rate κ_l. That is, if we have a measurement of ρ and a model of our system's evolution $\mathcal{L}_{\vec{\lambda}}$, we can optimize the state fidelity [Nielsen and Chuang, 2011]

$$\mathcal{F}(\rho, \rho(\vec{\lambda})) = \left(\mathrm{Tr}\left(\sqrt{\sqrt{\rho}\rho(\vec{\lambda})\sqrt{\rho}} \right) \right)^2 \tag{3.16}$$

over $\vec{\lambda}$ to find Hamiltonian parameters and loss rates κ_l.

But first, we need to find ρ. There are multiple approaches to quantum state tomography. See [Haroche and Raimond, 2006] for a collection of methods. Here, we will just explore one of these techniques. For the simple case of a qubit, the density

matrix can be expanded in terms of Pauli matrices

$$\rho = \frac{1}{2}\left(I + \langle X\rangle \sigma_X + \langle Y\rangle \sigma_Y + \langle Z\rangle \sigma_Z\right) \tag{3.17}$$

but no similar formula exists for bosonic cavity modes. Fortunately, there exists an invertible map between a measurable quantity and any density matrix, the Wigner function (or Wigner distribution). The Wigner function

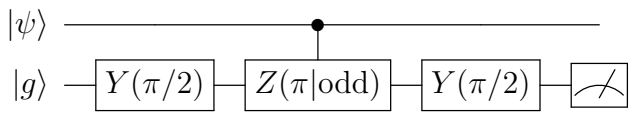
$$W(\alpha) = \frac{2}{\pi}\,\mathrm{Tr}\left(PD(-\alpha)\rho D(\alpha)\right) \tag{3.18}$$

is often called a "quasi-probability distribution" because it takes on real, but sometimes negative numbers [Haroche and Raimond, 2006]. It is a representation of the operator ρ in quantum phase space $\alpha \in \mathbb{C}$. Any operator on a Hilbert space can be mapped to a phase space representation via the Wigner-Weyl transform.

The operator P appearing in the definition of $W(\alpha)$ is the parity operator

$$P = e^{i\pi C^\dagger C} \tag{3.19}$$

with eigenvalue 1 for even Fock states and -1 for odd Fock states. We measure parity of a state $|\psi\rangle = \sum_n c_n |n\rangle$ using a Ramsey-like interference experiment



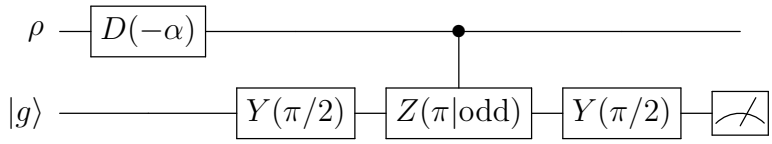
where the controlled Z in the circuit is a result of free evolution in the presence of

the dispersive interaction

$$e^{-iH_{\text{cQED}}t}\left(\frac{1}{\sqrt{2}}|\psi\rangle\otimes(|g\rangle+|e\rangle)\right)=\sum_n c_n|n\rangle\otimes\left(|g\rangle+e^{in\chi t_p}|e\rangle\right) \tag{3.20}$$

$$=\sum_n c_n|n\rangle\otimes(|g\rangle+(-1)^n|e\rangle) \tag{3.21}$$

and maps even photon number states to $|+X\rangle$ and odd to $|-X\rangle$ for $t_p=\pi/\chi$. We perform one final qubit rotation to measure in the X basis. From Eq. 3.18, we see that the Wigner function is a parity measurement of the displaced density matrix. We sample the Wigner function at each α



to build up the full distribution $W(\alpha)$. The left panel of Fig. 3.4 shows a measured Wigner function.

3.2.1 Finding ρ from $W(\alpha)$

Now that we have $W(\alpha)$, how do we get ρ? Let's revisit Eq. 3.18

$$W(\alpha)=\frac{2}{\pi}\text{Tr}\left(PD(-\alpha)\rho D(\alpha)\right) \tag{3.22}$$

$$=\frac{2}{\pi}\sum_{mn}\langle m|D(\alpha)PD(-\alpha)|n\rangle\,\rho_{nm} \tag{3.23}$$

$$=\sum_{mn}A_{mn}(\alpha)\rho_{nm} \tag{3.24}$$

and rewrite $A_{mn}(\alpha)$ and ρ_{mn} as vectors $\vec{A}(\alpha)$ and $\vec{\rho}$. The elements of $\vec{A}(\alpha)$ have a hideous form that we put below for reference [Cahill and Glauber, 1969]

$$A_{nm}(\alpha) = \frac{2}{\pi} e^{-2|\alpha|^2} \sqrt{\frac{m!}{n!}} (-1)^{m-1} (2\alpha)^{n-m} L_m^{n-m}(4|\alpha|^2). \tag{3.25}$$

This "vectorization trick" simply reorganizes indices (the order doesn't matter, as long as it's known and consistent) and will allow us to write a matrix equation relating a Wigner function vector to the vectorized density operator ρ. Our experimentally measured Wigner functions samples a discrete set of α_j

$$W(\alpha_j) = \vec{A}(\alpha_j) \cdot \vec{\rho} \tag{3.26}$$

where $\cdot$ is the vector dot product. This means that we can relate a vector of sampled Wigner points $(\vec{W})_j = W(\alpha_j)$ to $\vec{\rho}$ via

$$\vec{W} = A\vec{\rho} \tag{3.27}$$

with A promoted back to a matrix where $A_{ij} = \vec{A}_j(\alpha_i)$.

The simplest approach to solving Eq. 3.27 is to multiply both sides by the inverse A^{-1}, should A be an invertible matrix. Unfortunately, A is generally ill-conditioned with multiple singular values that make inversion unstable. This is usually solvable using the Moore-Penrose inverse $(A^\dagger A)^{-1} A^\dagger$ instead of A^{-1}. However, while $A^{-1}\vec{W}$ may formally be a valid solution for ρ, it may not produce a positive semidefinite Hermitian density operator that satisfies $\mathrm{Tr}(\rho) = 1$.

To find the density operator $\vec{\rho}$ that satisfies these physical requirements, we optimize $A\vec{\rho}$ to be the vector closest to the measured $\vec{W}$ by minimizing the Euclidean

norm

$$\min_{\vec{\rho}} |\vec{W} - A\vec{\rho}|_2 \tag{3.28}$$

using the convex optimization software CVXPY [Agrawal et al., 2018, Diamond and Boyd, 2016]. When performing this minimization, we must choose in advance the dimension of the solution ρ. We generally choose this number to be as small as possible to minimize any effects of ill-conditioning. The choice is made easy when we have some advance knowledge of the structure of ρ, such as a maximum photon number. In the case of Kerr characterization, ρ_0 is a coherent state whose photon number does not change. It is also important to normalize the measured Wigner function so that $\int W(\alpha) d^2\alpha = 1$. This condition corresponds to $\mathrm{Tr}(\rho) = 1$, which we are enforcing in the optimization. We show results of ρ reconstruction in Fig. 3.4.

It's worth noting at this point that the density operator reconstruction method finds the state closest to the *measured* Wigner function. If our Wigner function measurement has systematic errors, these errors are propagated into the reconstructed state ρ. For example, the $X(\pi/2)$ qubit rotation in the parity measurement circuit on page 45 is not purely unselective across all Fock states. The frequency content of the pulse supports an unselective rotation only over a finite number of Fock states. As the spread of Fock states comprising a state ρ increases, very high pulse bandwidths are required to maintain unselectivity (see Fig. 2.2).

If we are able to characterize the parity operator $\tilde{P}$ that our measurement actually implements, we can include its effect in the change-of-basis matrix A. Our new $\tilde{A}$ is

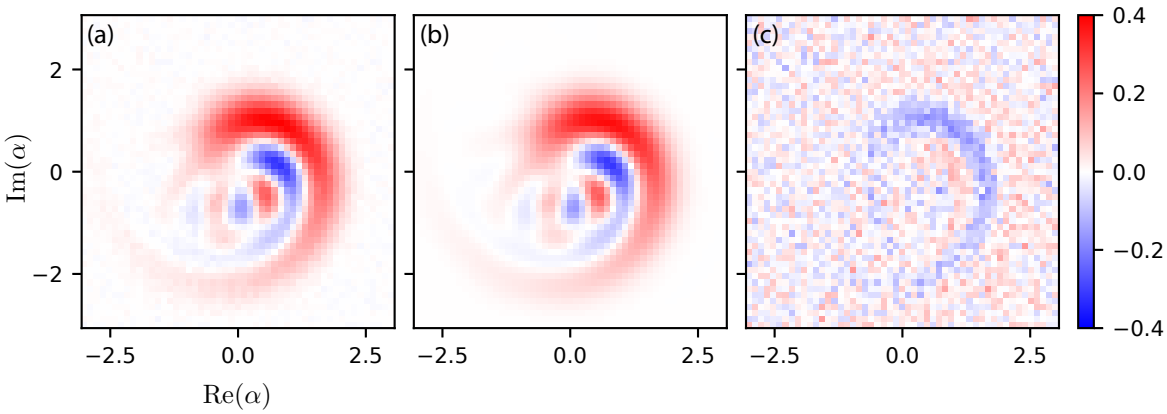


Figure 3.4: We prepared a coherent state $|\alpha = 1.5\rangle$ and allowed it to evolve for 10 μs under the Hamiltonian defined in Eq. 3.3. (a) We then measured the Wigner function, and reconstructed ρ using the optimization technique in Ch. 3.2.1. (b) For comparison, we computed the Wigner function of the reconstructed state ρ and (c) plotted the residuals (multiplied by a factor of 10).

defined as

$$W(\alpha) = \frac{2}{\pi} \operatorname{Tr}\left(\tilde{P} D(-\alpha) \rho D(\alpha)\right) \tag{3.29}$$

$$= \frac{2}{\pi} \sum_{mnjk} \langle m|D(\alpha)|j\rangle \, \tilde{P}_{jk} \, \langle k|D(-\alpha)|n\rangle \, \rho_{nm} \tag{3.30}$$

$$= \sum_{mn} \tilde{A}_{mn}(\alpha) \rho_{nm} \tag{3.31}$$

$$\tag{3.32}$$

where

$$\langle n|D(\alpha)|m\rangle = \begin{cases} e^{-|\alpha|^2/2} \sqrt{\frac{m!}{n!}} (\alpha)^{n-m} L_m^{n-m}(|\alpha|^2) \text{ for } n \geq m \\ e^{-|\alpha|^2/2} \sqrt{\frac{n!}{m!}} (-\alpha^*)^{m-n} L_n^{m-n}(|\alpha|^2) \text{ for } n \leq m. \end{cases} \tag{3.33}$$

as the sum over j, k is finite over the characterized dimension of $\tilde{P}$. This allows us to include the oft-observed contrast reduction in measurements of $W(\alpha)$ as $|\alpha|^2$ increases.

3.2.2 Finding K_C from fits of $W(\alpha)$ and ρ

Now that we have a reconstructed a density operator ρ, we can optimize the parametrized state $\rho(\vec{\lambda})$ from Eq. 3.14. For the purposes of Kerr characterization, we generate $\rho(\vec{\lambda})$ using

$$H(\vec{\lambda}) = \delta C^\dagger C - \frac{K_C}{2} C^\dagger C^\dagger C C + \frac{\beta}{6} \left(C^\dagger\right)^3 C^3 \tag{3.34}$$

along with cavity photon loss C with rate $\kappa_C = 1/(330\mu\text{s})$ for the system used in this chapter. We include a higher-order non-linearity as a diagnostic to monitor higher order effects amplified by the drive added to the transmon in the next section. We optimize over the three parameters $\vec{\lambda} = (\delta, K_C, \beta)$ in Eq. 3.34 using the cost function Eq. 3.16. We generate $\rho(\vec{\lambda})$ with master equation simulations in Qutip [Johansson et al., 2012, Johansson et al., 2013]. Results of this optimization are shown in Fig. 3.5.

We also perform a more direct optimization on the least-squares error between measured and simulated Wigner functions. This method cuts out the reconstruction step and is a good check that the reconstruction doesn't produce an unphysical state. The cost function for direct Wigner optimization is

$$\mathcal{F}_{\text{Wigner}} = \int (W(\vec{\lambda}, \alpha) - W_{\text{exp}}(\alpha))^2 d^2\alpha \tag{3.35}$$

where $W_{\text{exp}}(\alpha)$ is our measured Wigner function and $W(\vec{\lambda}, \alpha)$ is the Wigner function of $\rho(\vec{\lambda})$. Expanding the integral,

$$\mathcal{F}_{\text{Wigner}} = \int (W^2(\vec{\lambda}, \alpha) + W^2_{\text{exp}}(\alpha) - 2W(\vec{\lambda}, \alpha) W_{\text{exp}}(\alpha)) d^2\alpha \tag{3.36}$$

$$= \text{Tr}(\rho(\vec{\lambda})^2) + \text{Tr}(\rho^2) - 2\,\text{Tr}(\rho(\vec{\lambda})\rho) \tag{3.37}$$

we find that this cost function optimizes the trace overlap of ρ and $\rho(\vec{\lambda})$. Results of this fit are shown in Fig. 3.5 and closely match the fidelity optimization.

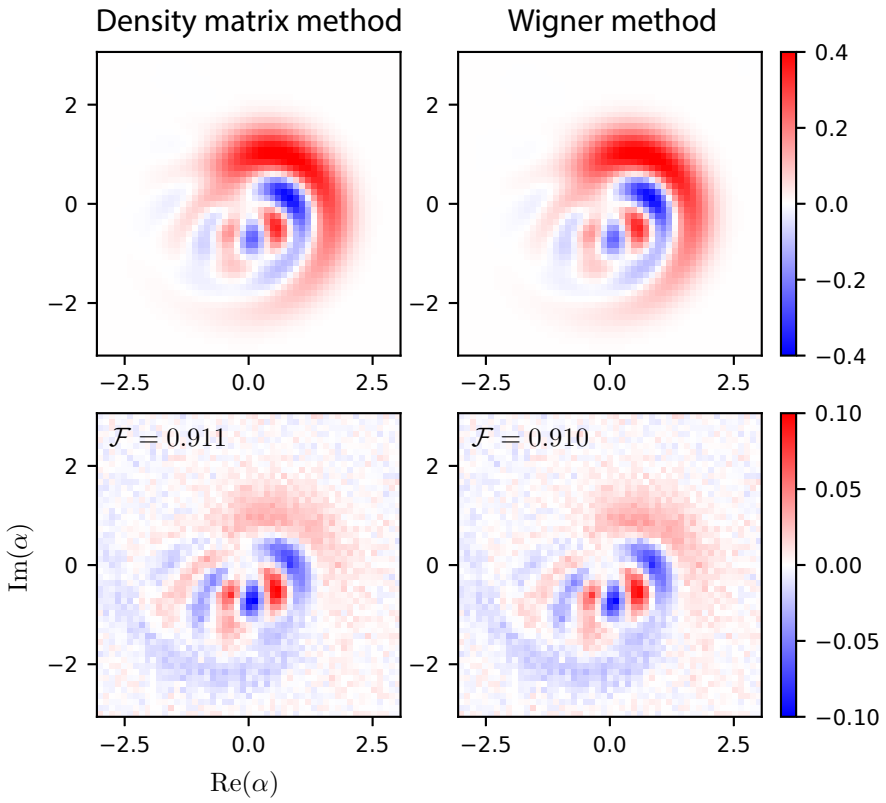


Figure 3.5: **Top left**: Wigner function of $\rho(\vec{\lambda})$ optimized against the reconstructed Wigner function shown in Fig. 3.4. The fidelity of the fit $\mathcal{F}(\rho, \rho(\vec{\lambda})) = 0.911$ is likely limited by transmon errors during the parity measurement. Residuals are shown **below** the Wigner function. The **right column** shows the same, but where $\rho(\vec{\lambda})$ was optimized using the Wigner fidelity in Eq. 3.35. The fidelity between the optimized states is 0.9995. Fit parameters $\vec{\lambda}$ are given in Table 3.1.

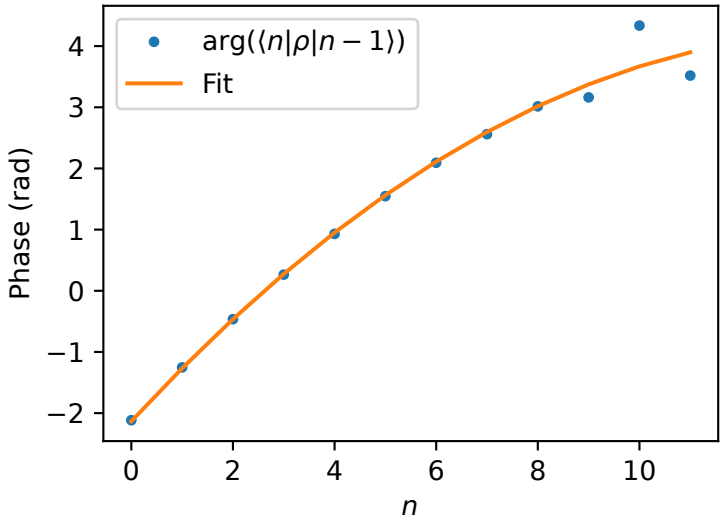


Figure 3.6: Fit of the phases $\arg(\langle n|\rho|n-1\rangle)$ to $(-\delta + K_C(n-1) - \frac{\beta}{2}(n-1)(n-2))t$ to find $\vec{\lambda}$. The reconstructed ρ here used is the same as in Figs. 3.4 and 3.5. Note that the fit becomes unreliable towards higher Fock states due to the initial state being a coherent state $|\alpha = 1.5\rangle$ with average photon number $\bar{n} = 2.25$. There are no statistical error bars available here. To perform error propagation through the density matrix reconstruction, we either need to employ bootstrapping or use autodifferentiation for error propagation. The fit results are in Table 3.1. Note that the value of the offset phase δt here is unreliable, since it can wrap through 2π.

There is one final way we will extract the parameters $\vec{\lambda}$ from ρ. If we assume that the idle evolution is purely unitary, then the Hamiltonian in Eq. 3.34 generates all of the evolution and

$$\rho(\vec{\lambda}) = e^{-iH(\vec{\lambda})t}\rho_0 e^{iH(\vec{\lambda})t}. \tag{3.38}$$

The phase of the off-diagonal elements evolves as

$$e^{-iH(\vec{\lambda})t}|n\rangle\langle n-1|e^{iH(\vec{\lambda})t} = e^{i(-\delta + K_C(n-1) - \frac{\beta}{2}(n-1)(n-2))t}|n\rangle\langle n-1| \tag{3.39}$$

meaning that we can fit the phases of the reconstructed density matrix $\langle n|\rho|n-1\rangle$ to extract $\vec{\lambda}$. This fit is performed in Fig. 3.6 with results in Table 3.1.

Table 3.1: Kerr characterization fit results for Figs. 3.6 and 3.5.

Method	δ	K_C	β	Fidelity $\mathcal{F}$
Density matrix fidelity fit	-84.65 kHz	-13.66 kHz	1.03 kHz	0.911
Wigner function fit	-84.28 kHz	-13.58 kHz	1.09 kHz	0.910
Off-diagonal element fit	N/A	-12.8 kHz	1.01 kHz	N/A

3.3 Kerr engineering

In general, the transmon-induced cavity Kerr in the bare Hamiltonian Eq. 2.19 is fixed and depends on the anharmonicity and participation of the transmon in the cavity mode. In this section, we describe and characterize a method for tuning cavity Kerr with a single weak drive on the transmon. The most complete derivation of this method can be found in [Zhang et al., 2022], but the method discussed there is beyond both the scope and skill of this dissertation. We derive the Kerr engineering method here using techniques more familiar to the experimentalist and then present experimental evidence of the technique. We include some start-up steps for others who may want to employ this technique.

3.3.1 Derivation and results

Consider a single pump on the transmon near the transmon frequency ω_A. Expanding Eq. 2.27 and neglecting terms rotating faster than ω_1 (all terms that don't preserve cavity photon number) produces an interaction term and pump-induced Stark shifts:

$$H = H_{\text{cQED}} - \chi \left(e^{i\omega_1 t} \varepsilon_1^* A C^\dagger C + \text{h.c.} \right) - 2K_A |\varepsilon_1|^2 A^\dagger A - 2K_C |\varepsilon_1|^2 C^\dagger C \tag{3.40}$$

In a frame rotating at ω_1, the full Hamiltonian becomes

$$H = (\omega_A - \omega_1)\, A^\dagger A + \omega_C C^\dagger C - \frac{\alpha}{2} A^\dagger A^\dagger A A - \frac{K_C}{2} C^\dagger C^\dagger C C \tag{3.41}$$
$$- \chi A^\dagger A C^\dagger C - 2\alpha|\varepsilon_1|^2 A^\dagger A - 2K_C|\varepsilon_1|^2 C^\dagger C - \underbrace{\chi\left(\varepsilon_1^* A C^\dagger C + \text{h.c.}\right)}_{V}$$

If $\chi|\epsilon_1|$ is small, we can treat the last term as a perturbation V. We are interested in finding corrections to $E_{|n,g\rangle}$ for states $|\text{cavity}, \text{qubit} = g\rangle$ as we want the transmon to stay in its ground state to prevent the dispersive interaction from dephasing the cavity. The first order correction from V is always zero

$$E^{(1)}_{|n,m\rangle} = \langle n, m|V|n, m\rangle = 0 \tag{3.42}$$

because $A, A^\dagger$ link adjacent levels. The second order correction $E^{(2)}_{|n,m\rangle}$ is non-zero

$$E^{(2)}_{|n,m\rangle} = \sum_{|\psi\rangle \neq |n,m\rangle} \frac{|\langle n, m|V|\psi\rangle|^2}{E_{n,m} - E_\psi} \tag{3.43}$$
$$= \frac{|\langle n, m|V|n, m+1\rangle|^2}{\omega_1 - \omega_A + n\chi + m\alpha - \chi' n^2/6} \tag{3.44}$$
$$= \frac{\chi^2|\varepsilon_1|^2 n^2(m+1)}{\omega_1 - \omega_A + n\chi + m\alpha - \chi' n^2/6} \tag{3.45}$$
$$= \frac{\chi^2|\varepsilon_1|^2 n^2(m+1)}{\omega_1 - \omega_A + m\alpha}\left(1 - \zeta(n) + O\left(\zeta(n)^2\right)\right) \tag{3.46}$$

for

$$\zeta(n) = \frac{n\chi - \chi' n^2/6}{\omega_1 - \omega_A + m\alpha} \tag{3.47}$$

which converges for Fock states $n < n_{\max}$ when $\zeta(n_{\max}) < 1$. This condition can be satisfied on a chosen manifold of states by choosing a sufficiently large pump detuning $\omega_1 - \omega_A$ (but not one so large that it activates other parametric processes!). We identify these energy shifts as matrix elements of an effective operator for the

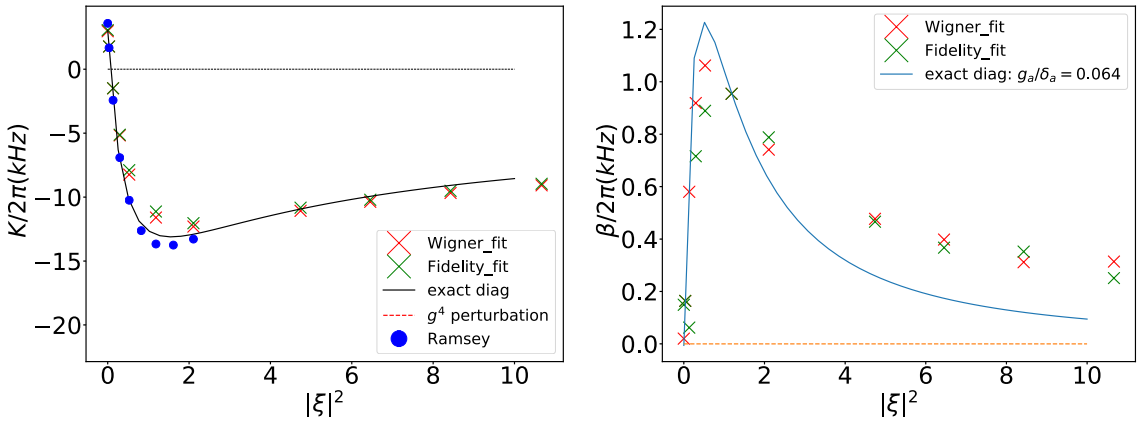


Figure 3.7: Kerr measurements and theory predictions with $\delta = 20$ MHz. "Exact diagonalization" refers to a numerical solution to Eq. 2.5 in the presence of a single pump. To acquire each point, we prepare $|\alpha = 1.5\rangle$, wait 10 μs in the presence of a drive with $\delta = 20$ and amplitude $|\varepsilon_1| = |\xi|$. The blue dots are data taken with the Ramsey experiments described in Section 3.1 and the two sets of crosses are produced via the Wigner fits described in Section 3.2.2. The fit of β (as defined in Eq. 3.34) is limited by $\alpha = 1.5$ which leaves little probability of more than two photons, reducing the effect of β. Note the small drive amplitude required to cancel Kerr in the cavity. The induced β at this point is also very small.

perturbation V

$$V_{\text{eff.}} = \frac{\chi^2|\varepsilon_1|^2}{\omega_1 - \omega_A + m\alpha} C^\dagger C^\dagger C \left(m|m\rangle\langle m| + 1\right) \tag{3.48}$$

to first order in $1/(\omega_1 - \omega_A + m\alpha)$, which looks similar to the non-linear dispersive interaction χ'. The $O(1/(\omega_1 - \omega_A + m\alpha)^2)$ term in Eq. 3.46 modifies the distortion arising from higher-order nonlinearities like sixth-order cosine term $\frac{\beta}{6}\left(C^\dagger\right)^3 C^3$, but is suppressed by an additional order of the pump detuning. That is why we added this non-linearity to our fits in Ch. 3.2.1.

Since we are interested in the effect of this pump when the qubit is in the ground state only, we construct an effective Hamiltonian for Eq. 3.41 to first order in $1/(\omega_1 -$

$\omega_A + m\alpha$) by replacing $V \to \langle g|V_{\text{eff.}}|g\rangle$:

$$H_{\text{eff.}} = \left(\omega_A - \omega_1 - \alpha|\varepsilon_1|^2\right) A^\dagger A + \left(\omega_C - \frac{\chi^2|\varepsilon_1|^2}{\omega_1 - \omega_A} - 2K_C|\varepsilon_1|^2\right) C^\dagger C \tag{3.49}$$

$$- \frac{\alpha}{2} A^\dagger A^\dagger A A + \left(\frac{\chi^2|\varepsilon_1|^2}{\omega_1 - \omega_A} - \frac{K_C}{2}\right) C^\dagger C^\dagger C C. \tag{3.50}$$

Note that the sign of the detuning $\omega_1 - \omega_A$ determines the sign of the added effective Kerr. Cavity Kerr is canceled when

$$\frac{\chi^2|\varepsilon_1|^2}{\omega_1 - \omega_A} - \frac{K_C}{2} = 0 \Rightarrow |\varepsilon_1|^2 = \frac{\omega_1 - \omega_A}{4\alpha} \tag{3.51}$$

which generally only requires small pulse amplitudes. Second-order perturbation theory only applies when V is small, requiring other methods to predict the effect of the drive. Exact numerical diagonalizations provide the most accurate prediction, and agree with experimental results for K_C, β at different drive powers $|\varepsilon_1|^2 = |\xi|^2$, as shown in Fig. 3.7.

One concern when adding a drive to the transmon is that we may activate higher-order processes that damage system coherences [Zhang et al., 2019]. We show in Fig. 3.8 that the qubit and cavity lifetimes are only weakly dependent on the drive, but that the transmon heating rate doubles at high amplitude. The deterministic Stark shift caused by the pump can be accounted for by updating the phase of the cavity local oscillator at the end or during the pump.

3.3.2 Tune-up procedure

Tuning up the Kerr-cancelling drive doesn't generally require fully characterizing drive strengths ϵ_1 and higher-order nonlinearities β. Since the amplitude required to cancel Kerr is generally small, we completely neglect β and recommend the following tuneup procedure:

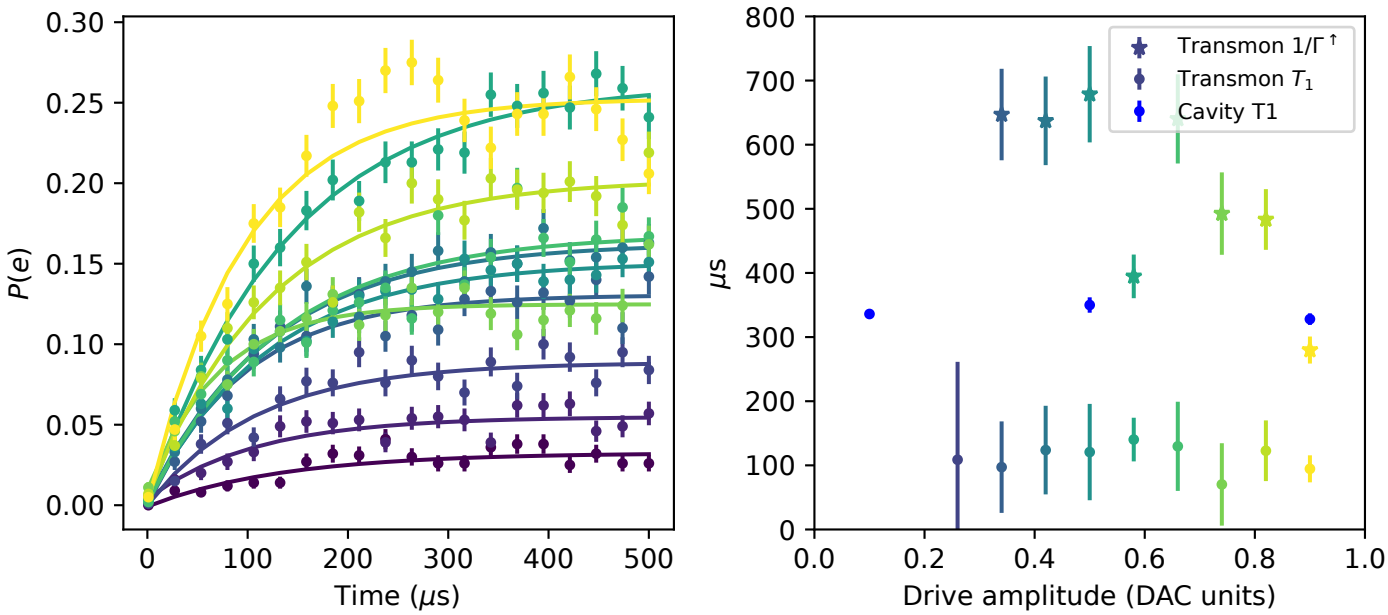


Figure 3.8: **Left**: transmon excited state populations for different Kerr engineering pump amplitudes applied for a scanned duration with $\delta = 20$ MHz. The color scheme for both panes corresponds to the drive amplitude, which can be matched to the x-axis in the right pane. Each trace is fit to an exponential to find $\Gamma^{\uparrow}$ and $\Gamma^{\downarrow}$. The equilibrium excited state populations are plateaus of each fit. **Right**: cavity T_1s at three different drive amplitudes. The other dots are transmon $1/\Gamma^{\downarrow}$ and stars are transmon $1/\Gamma^{\uparrow}$. Note that the results of the smallest two amplitudes (0.1 and 0.18) have low excited state populations and are left off of the right plot due to resulting high uncertainties in the fit. The amplitude required to cancel K_C is typically less than 0.1, where the transmon equilibrium population remains below 3%.

1. Choose a drive detuning first. We recommend using a blue-detuned drive so that Stark shifts don't push the qubit closer to the drive. 20-50 MHz blue detuned from the transmon is generally a good starting point.

2. Verify that the drive doesn't heat the qubit by applying the pump with the transmon in the ground state and scanning over amplitude and duration. Use the results to replicate Fig. 3.8 and learn a safe range of amplitudes. If you are tuning up the Kerr cancelling pump to operate during another parametric process (such as squeezing), detune the other drives from their resonance conditions. All drives should be applied at this step to detect any unwanted heating or parametric effects before we continue.

3. Measure Kerr without the Kerr cancelling pump (add off-resonant squeezing or beamsplitter drives if applicable). We recommend using the Ramsey method in Section 3.1 using coherent states for the initial Kerr tuneup, and Fock states afterwards for higher accuracy.

4. Add the Kerr cancelling pump and sweep the amplitudes to find a zero crossing. If tuning up with other parametric drives, tweak those frequencies to zero out the Stark shift from all drives. If not tuning up with other drives, note the Stark shift and update the local oscillator phase accordingly. This update should be performed during the drive so as not to add an additional unpumped delay at the end.

Chapter 4

Bitwise measurement

In the previous chapter, we used end-of-line (EOL) measurements to characterize the strength of cavity Kerr and tune its cancellation. This chapter combines EOL and QND measurements to explore the bitwise measurement first utilized in a simulation of Franck-Condon factors [Wang et al., 2020]. With the exception of Sec. 4.2, this chapter is reprinted with permission from [Curtis et al., 2021], copyright 2021 by the American Physical Society.

4.1 Introduction

EOL measurements in cQED extract one bit of information per transmon readout corresponding to the current state of the transmon qubit. This type of single-bit measurement is common in many quantum systems. For example, in photonic platforms, the analog of transmon measurement is single photon detection. Single photon detectors (SPDs) distinguish between zero and at least one photon, and are essential for tasks such as measuring the output of a photonic register or heralding in probabilistic quantum gates. Uses of SPDs include quantum key distribution [Lo et al., 2012, Yin et al., 2016, Liao et al., 2017], linear optical quantum computing (LOQC) [Knill et al., 2001, Kok et al., 2007], quantum communication [Duan et al., 2001, Ursin

et al., 2007, Kimble, 2008, Northup and Blatt, 2014, Hu et al., 2016, Zhang et al., 2017, Dou et al., 2018], and photonic quantum simulations [Spring et al., 2013, Tillmann et al., 2013, Sparrow et al., 2018, Bentivegna et al., 2015]. Highly efficient SPDs with low dark count rates based on avalanche photodiodes and, more recently, superconducting nanowire single-photon detectors [Eisaman et al., 2011, Marsili et al., 2013, Esmaeil Zadeh et al., 2017] have been developed and satisfy the demands of these optical applications.

The information processing capacity of intrinsically bosonic optical modes, however, can be much greater by executing protocols that manipulate multiphoton states. In these cases, SPDs are insufficient for distinguishing between multiple photons and photon number-resolving (PNR) detectors are required instead. Proposals to enhance quantum communication [Simon et al., 2007] and key distribution [Cattaneo et al., 2018] protocols, perform teleported gates [Knill et al., 2001], and extend conventional boson sampling [Spring et al., 2013, Huh et al., 2015, Hamilton et al., 2017, Clements et al., 2018, Kruse et al., 2019] all require PNR detectors. Building an optical PNR detector has proven to be a difficult task. There are a variety of promising approaches relying on highly efficient SPDs in multiplexing or arraying schemes [Banaszek and Walmsley, 2003, Fitch et al., 2003, Achilles et al., 2004, Divochiy et al., 2008, Mattioli et al., 2016, Tao et al., 2019] and transition edge sensors [Kardynał et al., 2008, Gerrits et al., 2011, Calkins et al., 2013]. These recent advances have vastly improved performance, but the limited fidelity and resolvable photon number of optical PNR detectors hamper multiphoton experiments [Clements et al., 2018].

The detection of single microwave photons is less established and more challenging, due to their lower energy and higher thermal background levels. The dispersive interaction between a photon and an atom or qubit enables a wide range of measurement capabilities not possible in the optical domain. In cavity QED, the dispersive interaction with Rydberg atoms enabled the observation of single pho-

ton jumps [Gleyzes et al., 2007]. In cQED systems, Josephson junctions coupled to microwave cavities generate the dispersive interaction essential for measuring single microwave photons [Schuster et al., 2007, Johnson et al., 2010]. Additionally, the dispersive interaction has been used to create [Houck et al., 2007, Hofheinz et al., 2008] and manipulate [Heeres et al., 2015, Heeres et al., 2017] these photons in a wide variety of ways. cQED analogs of optical photodetectors such as the Josephson photomultiplier [Chen et al., 2011, Govia et al., 2014, Opremcak et al., 2018] have also been developed.

Introducing high-performance PNR detectors to the microwave regime could greatly enhance the prospects for performing boson sampling and other photonic quantum information processing techniques with cQED systems. Early quantum non-demolition (QND), single-shot, number-resolving measurements of microwave photons used flying Rydberg atoms to iteratively update the experimentalists' knowledge of the photon number distribution [Guerlin et al., 2007]. Previous efforts in cQED to develop PNR measurements have fallen short of combining single-shot and QND capabilities. Spectral density analysis [Hofheinz et al., 2008] and state interrogation methods [Schuster et al., 2007] are QND, but require many shots of the experiment to build up time or spectroscopic traces. Parity measurements have been used to count photon jumps [Sun et al., 2014], but, while single-shot, the full PNR measurement is not QND. More recently, frequency multiplexing [Essig et al., 2021] and multiple feedforward measurements [Dassonneville et al., 2020] have been used to make QND, single-shot PNR measurements, but suffer from limited fidelity.

In this chapter, we describe a measurement protocol that implements a high-fidelity, single-shot PNR detector of microwave photons. The single-shot nature of this measurement is essential to sample from the exponentially large Hilbert space inherent to experiments with multiple bosonic modes. We previously introduced this protocol to efficiently sample from the probability distribution at the output of a

multiphoton bosonic quantum simulator, reducing the required number of measurements by a factor of 256, the size of the Hilbert space [Wang et al., 2020]. Errors in the storage and ancilla modes caused non-trivial bias in the output, but these errors are well understood and can be modeled with a simple hidden Markov model. We calibrate and use this model to improve the fidelity of the measurement by an order of magnitude using error mitigation methods, reducing the measurement infidelity to a few percent.

This chapter is structured as follows. First, in Section 4.2, we discuss an application for which we devised our new single-shot PNR measurement. Then in Section 4.3, we discuss QND measurements of multiple observables with a single prepared state. We verify measurements of parity and its generalizations satisfy the conditions required to be QND. Next, we show how to construct a PNR detector of microwave photons from successive QND measurements of parity and its generalizations. This is accomplished by representing the photon number in its binary decomposition. In Section 4.5, we introduce the error syndromes of our system along with a hidden Markov model (HMM) that parametrizes the errors. After calibrating the error rates, we use the model to mitigate the errors on an ensemble of states using deconvolution methods in Section 4.5.4. Finally, we show that our error deconvolution protocol is scalable in the number of modes and maintains computational efficiency.

4.2 Motivation: Simulation of Franck-Condon Factors

Before detailing the implementation of our single-shot PNR detector, we provide an example application as motivation [Wang et al., 2020]. As mentioned above, bosonic systems offer an exciting platform for quantum simulation problems. One such problem that seems to have a computational advantage is Gaussian boson sampling [Kruse

et al., 2019,Hamilton et al., 2017,Bentivegna et al., 2015,Lund et al., 2014] which has a practical use in simulating Franck-Condon factors (FCFs) [Huh et al., 2015,Clements et al., 2018,Huh et al., 2020].

One focus of quantum simulation so far has been predicting the electronic structure of molecules [Whitfield et al., 2011,Kandala et al., 2017,Kandala et al., 2019,Kivlichan et al., 2018]. Exact classical computations of molecular electronic structure are computationally difficult, requiring a zoo of complicated approximations, algorithms, and computational methods [Fales et al., 2020] [1] to extract quantities useful to chemists.

One such approximation, the adiabatic Born-Oppenheimer approximation, allows separation of the molecular wavefunction into electronic and vibrational components. The electronic component describes the occupation of molecular orbitals while the vibrational component describes quantum behavior of the positions of atomic nuclei comprising the molecule. The vibrational wavefunctions are determined by a Hamiltonian that is a function of the electronic state that determines the distribution of electrical charge. When a change in the electronic state occurs, the Hamiltonian of the vibrational wavefunction changes, producing changes in the vibrational state. These transitions are called "vibronic" transitions, whose probability amplitudes can be computed with Gaussian boson sampling.

When one expands the vibrational potential energies about their equilibrium points, we can approximate the vibrational Hamiltonian as a collection of harmonic oscillators, which have exactly the same structure as our superconducting cavity oscillators. The deformation of the potential energy that occurs during a vibronic transition changes the equilibrium positions, energies, and hybridization of this collection of harmonic oscillators. We express the deformation of the mode operators $\vec{a}$

1. Disclaimer: one of the co-authors on this publication is the author's brother

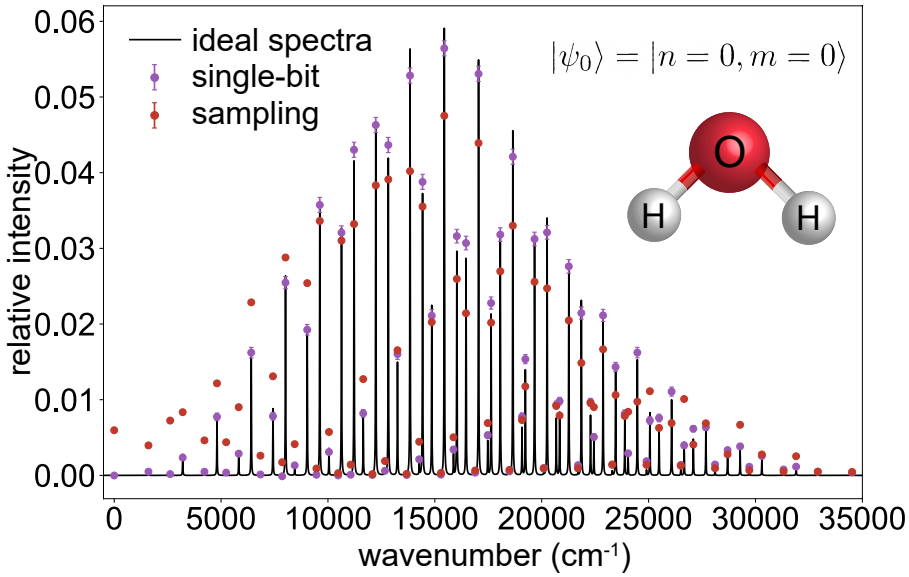


Figure 4.1: Example Franck-Condon factor simulations from [Wang et al., 2020]. The vibrational modes of water can approximated with two harmonic oscillators. In this figure, we plot transition probabilities from the ground vibrational state $|\psi_0\rangle = |n = 0, m = 0\rangle$ as a function of final state $|n', m'\rangle$ wavenumber $\nu = n'\nu_1 + m'\nu_2$. The single-bit measurement is clearly more accurate, with the "bitwise" sampling measurement producing substantial discrepancies at some wavenumbers, but less at others. The error patterns and how to mitigate them are discussed in Section 4.5.

with the Duschinsky transformation [Duschinsky, 1937]

$$\vec{a} \to U_{\text{Dok}} \vec{a} U_{\text{Dok}}^{\dagger} \tag{4.1}$$

in its Doktorov decomposition [Doktorov et al., 1977] of Gaussian operators

$$U_{\text{Dok}} = D(\vec{\alpha}) S^{\dagger}(\vec{\xi}') R(U) S(\vec{\xi}) \tag{4.2}$$

where D and S are the tensor product of single-mode displacement and squeezing operators. $R(U)$ is a N-mode rotation operator that can be decomposed into two mode beamsplitters. cQED implementations of each of these Gaussian operations exist via the Hamiltonian engineering methods discussed in Section 2.2.1. For more details, see [Wang et al., 2020, Wang, 2022].

The transition probabilities (FCFs) are

$$\text{FCF}|_{\vec{n},\vec{n}'} = |\langle \vec{n}'|U_{\text{Dok}}|\vec{n}\rangle|^2 \tag{4.3}$$

where $|\vec{n}\rangle = |n_1\rangle \otimes |n_2\rangle \otimes \ldots |n_N\rangle$ is a tensor product of Fock states with energy of $|\vec{n}\rangle$ is $\sum_k \omega_k n_k$. We simulate the vibronic transition by preparing $|\vec{n}\rangle$ and applying U_{Dok}. Finding the probabilities $\text{FCF}|_{\vec{n},\vec{n}'}$ is then reduced to sampling this state to build up a probability distribution as a function of $|\vec{n}'\rangle$. There are two ways to go about this sampling:

1. Extract a single bit of information per shot by answering the question "is the quantum state $|\vec{n}'\rangle$?" This type of measurement can be implemented with the number-selective π-pulses discussed in 2.2.1 with each transmon in the N-mode system checking if the cavity to which it's coupled has n'_k photons. Suppose that we want to measure up to N_{max} photons per mode. Then the number of times we have to ask the question "is the quantum state $|\vec{n}'\rangle$?" is N_{max}^N. Any scheme that requires an exponential number of measurements is clearly not scalable, so we need another approach. Even for $N = 2$ and $N_{\text{max}} = 16$ simulations we performed in [Wang et al., 2020], this requires 256 measuremments.

2. Extract the number of photons present in each cavity mode per shot. For up to N_{max} photons per mode, this requires measuring $\log_2(N_{\text{max}})$ bits per mode. To extract multiple bits of information in a single-shot of the experiment, we need a special type of measurement that does not perturb the state, so that more information can be extracted. We discuss the principle of operation of this measurement and demonstrate its viability in this chapter. Crucially, this measurement solves the exponential sampling problem as the total number of measurements $N_{\text{max}}N$ is now linear in the number of modes. The number of measurements per mode remains constant even as more modes are added. In

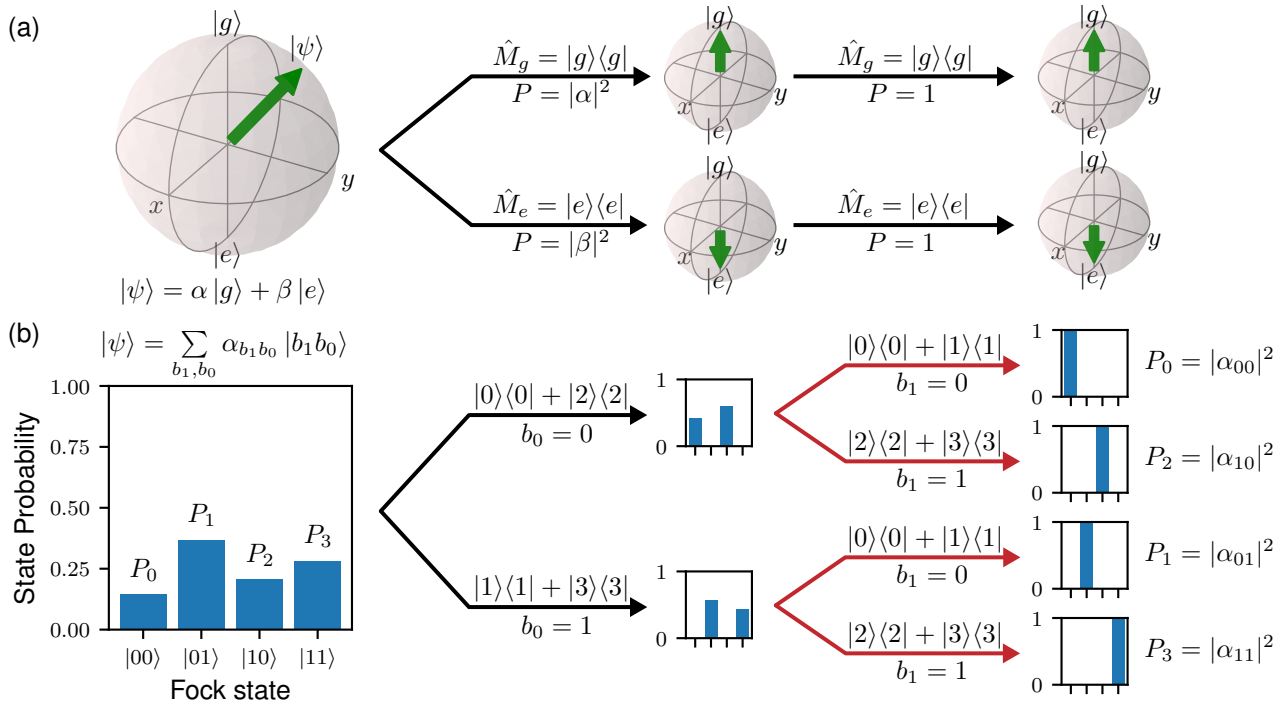


Figure 4.2: Projective, binary-valued measurements. (a) Measuring a qubit in its energy eigenbasis (the basis of σ_z eigenstates) projects the qubit into its ground state $|g\rangle$ with probability $|\alpha|^2$ or its excited state $|e\rangle$ with probability $|\beta|^2$. The measurement operators $\hat{M}_g$ and $\hat{M}_e$ commute with the Hamiltonian at all times. As a result, the shared eigenstates are not perturbed by additional measurements or time evolution, making the measurement QND. (b) Here, Fock states are represented with the first two bits (parity and super-parity) of their binary decomposition $|n\rangle = |b_1(n)b_0(n)\rangle$ for $N < 4$. An initial measurement of parity (black arrows labeled by measurement outcome and corresponding measurement operator) projects $|\psi\rangle$ into the even or odd subspace, and renormalizes the remaining state probabilities. A subsequent measurement of super-parity (red arrows), further projects the state fully into a single Fock state $|n\rangle$. The detected photon number is computed using a record of each measurement.

Fig. 4.1, we show simulated FCFs for water using both the single-bit measurement and the bitwise measurement. The single-bit measurement appears more accurate than the bitwise measurement, whose errors have a more complicated pattern that we address in Section 4.5.

4.3 Background

Optical photon detectors such as photomultiplier tubes, avalanche photodiodes, and superconducting nanowire SPDs function by converting incoming photons into electrical charges which are then amplified to produce a detectable signal [Eisaman et al., 2011]. This process is single-shot, but is inherently destructive as the incoming photons are consumed by the detector. Standard techniques in cQED for measuring single photons are QND [Johnson et al., 2010, Sun et al., 2014], allowing for additional processing of the state. Our protocol for PNR detection utilizes this ability to perform multiple measurements that collectively form a number-resolving measurement. In this section, we generalize the notions of QND and single-shot to a multi-level system with multiple measurements.

4.3.1 Multi-level QND measurements

To realize a single-shot PNR detector, we need to extend the concept of QND measurements to the case of a multi-level system and multiple measurements. Not all observables can be measured in a QND way; they must satisfy two conditions [Haroche and Raimond, 2006]. The first QND condition is that the observable $\hat{\mathcal{A}}$ commutes with the system Hamiltonian $\hat{H}$ to ensure that a repeated measurement of it yields the same result at any subsequent time (in the absence of external perturbations, such as photon loss). The second condition is that $\hat{\mathcal{A}}$ commutes with the measurement Hamiltonian $\hat{H}_M$, which contains additional terms that couple to a meter. This condition protects the state from additional evolution induced by the measurement process. QND measurements implement measurement operators $\hat{M}_\lambda$ for each outcome λ, which project the system into the eigenstate labeled by its eigenvalue λ.

Let us first consider the usual case of QND measurements of a qubit or two-level system initialized in an arbitrary state $|\psi\rangle = \alpha\,|g\rangle + \beta\,|e\rangle$. A measurement of the

system in this basis, with measurement operators $\hat{M}_g = |g\rangle\langle g|$ and $\hat{M}_e = |e\rangle\langle e|$, projects $|\psi\rangle$ into $|g\rangle$ with probability $|\alpha|^2$ and $|e\rangle$ with probability $|\beta|^2$. The outcome of this measurement tells us exactly the state of the qubit following the measurement. If the measurement of the observable is QND, then repeated measurements at any subsequent time will yield the same result, as seen in Fig. 4.2 (a).

How can we extend this concept to perform a QND measurement on a multilevel system? For instance, consider a four-level subspace of a bosonic mode C with operators $C, C^\dagger$. Two bits of classical information must be obtained to determine the state of this system. This can be accomplished with two QND measurements, whose composition projects the system into a single level. The first QND measurement halves the Hilbert space, and projects the state into one of its two subspaces, extracting one bit of classical information. A second measurement further halves this subspace, extracting the second bit and projecting the state into a single level of s. This is only possible if the two operators commute, else the second measurement would project into a superposition of states from the two subspaces of the first measurement, spoiling the first bit of information.

For example, the binary decomposition of the number of excitations in s contains two bits identifying the state of the four level system discussed above. Parity $\hat{\mathcal{P}}^0$ divides the space into even and odd subspaces with measurement operators $\hat{B}^{(0)}_{b_0=0} = |0\rangle\langle 0| + |2\rangle\langle 2|$ and $\hat{B}^{(0)}_{b_0=1} = |1\rangle\langle 1| + |3\rangle\langle 3|$. This measurement projects the system into the state corresponding to measurement outcome b_0, which is the least significant bit in the binary decomposition of the excitation number of s. Measurements of super-parity $\hat{\mathcal{P}}^1$, which commutes with $\hat{\mathcal{P}}^0$, measure the second-least-significant bit b_1 with measurement operators $\hat{B}^{(1)}_{b_1=0} = |0\rangle\langle 0| + |1\rangle\langle 1|$ and $\hat{B}^{(1)}_{b_1=1} = |2\rangle\langle 2| + |3\rangle\langle 3|$. Measured in succession, these two observables project the system into a single Fock state labeled by its binary decomposition $|b_1 b_0\rangle$, as shown in Fig. 4.2 (b). These generalized parity measurements form the basis of our approach to realize a QND, PNR detector of

microwave photons in a cQED system.

4.3.2 Multi-level single-shot measurements

As we discuss in the previous section, measuring a two-level or multi-level system requires extracting one or more bits of information. For these measurements to be considered single-shot, they must accurately measure a significant majority of their bits in a single shot. When measuring a qubit, this means that one gains close to one bit of information about the state per measurement, per initial state subjected to that measurement. With a correspondingly high measurement efficiency and a low dark count rate, a qubit measurement then extracts enough of the available information that we can measure it with precision close to the shot noise limit imposed by the ensemble size.

We call a multi-level measurement single-shot if a significant fraction of the bits are faithfully extracted per shot. For example, choosing 3/4 as this fraction corresponds to extracting more than three bits of information per shot for a PNR detector with $N = 16$ levels. To determine $P(n)$ of a multi-level system, we must sample many times from an ensemble of states. If the PNR measurement that acquires each sample meets this threshold, the error in the sampled $P(n)$ is limited by the shot noise of the number of samples. This capability is particularly crucial to extract joint probability distributions from M N-level systems without individually checking each of the N^M states. We discuss the single-shot character of our detector in Section 4.5.2.

4.4 The bitwise measurement

4.4.1 Measuring binary-valued cavity observables

Our PNR detector uses a qubit ancilla to measure binary-valued observables of a bosonic storage mode via the dispersive interaction. These observables $\hat{\mathcal{O}}$ divide the

Hilbert space into exactly two subspaces with distinct eigenvalues λ_1, λ_2. We can imagine a gate that entangles the two ancilla states with the eigenspaces $E_{\lambda_1}, E_{\lambda_2}$. A readout of the ancilla would then project the storage mode into the corresponding eigenspace. This two-step measurement is QND only if the QND conditions are satisfied for all steps in the measurement. For a measurement of $\hat{\mathcal{O}}$ to satisfy the first QND condition, it must be diagonal in the energy eigenbasis.

We next verify that the second QND condition is satisfied. Our three-mode cQED system consisting of a readout mode, a storage mode, and a transmon ancilla has operators $\hat{r}, \hat{r}^\dagger$, $\hat{C}, \hat{C}^\dagger$, and $\hat{\sigma}_z$, respectively. The composite system Hamiltonian is

$$\hat{H}/\hbar = \frac{\omega_\sigma}{2}\hat{\sigma}_z + \omega_s \hat{C}^\dagger \hat{C} + \omega_r \hat{r}^\dagger \hat{r} - \frac{\chi_{C\sigma}}{2}\hat{C}^\dagger \hat{C}\hat{\sigma}_z - \frac{\chi_{r\sigma}}{2}\hat{r}^\dagger \hat{r}\hat{\sigma}_z. \tag{4.4}$$

The device is the same as used in Refs. [Elder et al., 2020, Reinhold et al., 2020] with similar parameters, unless otherwise noted. We first apply an ancilla-storage entangling drive, which produces the mapping Hamiltonian $\hat{H}_{\text{map}}/\hbar = \hat{H}/\hbar + \Omega(t)\hat{\sigma}_x$. For observables $\hat{\mathcal{O}}$, frequency-selective drives $\Omega(t)$ can implement the requisite entangling operation by exciting the ancilla conditioned on a Fock state. The second step, ancilla readout, also utilizes a dispersive interaction, realized with the Hamiltonian $\hat{H}_{\text{readout}}/\hbar = \hat{H}/\hbar + \epsilon(t)\hat{r} + \epsilon^*(t)\hat{r}^\dagger$. This readout completes the measurement of $\hat{\mathcal{O}}$. Both the mapping and readout Hamiltonians commute with the diagonal $\hat{\mathcal{O}}$, and satisfy the second QND condition. These drives can induce dephasing [Reinhold et al., 2020], but do not induce additional decay in the storage mode [Sun et al., 2014]. Measurements of $\hat{\mathcal{O}}$ are thus QND, enabling the measurement of multiple generalized parity operators that form the basis of our number-resolving measurement.

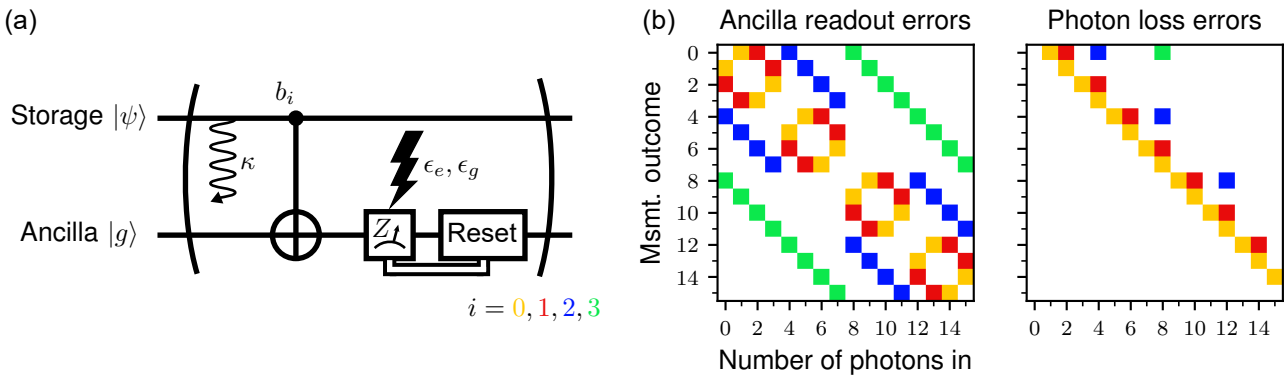


Figure 4.3: Measurement circuit diagram and error syndromes. (a) The photon number of a state $|\psi\rangle = \sum_n c_n |n\rangle$ in the storage cavity with decay rate κ is measured by sequentially interrogating the least-significant to most-significant bit in the binary decomposition of the photon number ($i = 0, 1, 2, 3$). Optimal control pulses excite the ancilla conditioned on the value of b_i in the storage cavity, followed by a dynamic reset of the ancilla to ensure that it starts in its ground state for the subsequent bit measurement [Elder et al., 2020] (see Appendix A.1 for reset statistics). The combined map-measurement process has associated error rates $\epsilon_{e(g)}$ for mis-assigning the ancilla to be in the $|g\rangle$ ($|e\rangle$) state when it should be $|e\rangle$ ($|g\rangle$). (b) Single ancilla (left panel) and decay (right panel) errors associated with the different bits (0 – yellow, 1 – red, 2 – blue, and 3 – green) produce different, but partially overlapping, error syndromes. Note that the syndromes for the decay errors depend on the order in which the bits are measured, and that the entries are qualitative labels of the measurement errors occurred and are not indicative of error magnitudes.

4.4.2 Generalized parity measurements

To implement the number-resolving measurement, we synthesize gates that map the binary-valued generalized parity operators onto the ancilla qubit. Each gate enables the measurement of one bit, so to resolve the first N states, we require $B = \log_2 N$ gates. The generalized parity operators

$$\left(\hat{\mathcal{P}}^k\right)_{ij} = \begin{cases} 0 & \text{for } i \neq j \\ 1 - 2\left(\left\lfloor \frac{i}{2^k} \right\rfloor \pmod 2\right) & \text{for } i = j \end{cases} \tag{4.5}$$

with eigenvalues $\lambda = \pm 1$ halve the Hilbert space into two eigenspaces $E_{+1} = \{n | b_k(n) = 0\}$ and $E_{-1} = \{n | b_k(n) = 1\}$, where $b_k(n)$ is the k^{th} bit in the binary decomposition of n. The $\hat{\mathcal{P}}^k$ are diagonal in the energy eigenbasis and thus satisfy both conditions for QND measurement. To perform this measurement, we use numerical optimal control techniques (see Appendix A.3 for more details) to synthesize a CNOT-like unitary operation

$$\begin{aligned} \mathrm{C}\hat{\mathcal{P}}^k = & \sum_{b_k(n)=0} e^{i\varphi_n} |g\rangle\langle g| \otimes |n\rangle\langle n| \\ & + \sum_{b_k(n)=1} e^{i\varphi_n} |e\rangle\langle g| \otimes |n\rangle\langle n| \\ & + \sum_{b_k(n)=1} e^{i\varphi'_n} |g\rangle\langle e| \otimes |n\rangle\langle n|. \end{aligned} \tag{4.6}$$

The gate imparts phases φ_n, φ'_n, but these do not affect the final measurement outcome because the composite measurement projects into a single Fock state, rendering the induced phase irrelevant. We can set these phases ourselves, but here allow the optimizer flexibility to choose the phases. Applying this gate to an ancilla prepared in $|g\rangle$ coupled to a storage mode with state $|\phi\rangle = \sum_n c_n |n\rangle$ entangles the odd and

even subspaces with the state of the ancilla,

$$\begin{aligned} \mathrm{C}\hat{\mathcal{P}}^k(|g\rangle \otimes |\phi\rangle) = & \sum_{b_k(n)=0} c_n e^{i\varphi_n} |g\rangle \otimes |n\rangle \\ & + \sum_{b_k(n)=1} c_n e^{i\varphi_n} |e\rangle \otimes |n\rangle \,. \end{aligned} \tag{4.7}$$

QND measurements of the ancilla then project the storage state into either E_{+1} or E_{-1} depending on the outcome, realizing the measurement operators

$$\hat{B}^{(k)}_{0(1)} = \sum_{b_k(n)=0(1)} e^{i\varphi_n} |n\rangle\langle n|. \tag{4.8}$$

The act of measuring B parity operators $\hat{\mathcal{P}}^{B-1}, \ldots, \hat{\mathcal{P}}^1, \hat{\mathcal{P}}^0$ with outcomes $b_{B-1}, \ldots, b_1, b_0$ in the subspace including only the first N states is equivalent to applying the measurement operator

$$\hat{M}_n = \hat{B}^{(B-1)}_{b_{B-1}} \ldots \hat{B}^{(1)}_{b_1} \hat{B}^{(0)}_{b_0} = |n\rangle\langle n| \tag{4.9}$$

and renormalizing. This operator projects the system into the Fock state $|n\rangle = |b_{B-1}(n) \ldots b_1(n) b_0(n)\rangle$ completing the "bitwise" measurement. This approach requires the minimal number of binary measurements of any scheme assuming no prior knowledge of the state. This protocol implements a detector resolving up to 2^B photons, where the number of generalized parity (bit) measurements we make is fully programmable.

4.4.3 Experimental implementation

To resolve the storage mode photon number in a single shot, we sequentially measure the generalized parity operators, whose outcomes form the binary decomposition of the photon number. For each bit, we apply $\mathrm{C}\hat{\mathcal{P}}^k$, read out the ancilla, and dynamically reset it using the same method as in Ref. [Elder et al., 2020]. More details of the reset

protocol, including statistics on the number of attempts required, are in Appendix A.1. This process is concatenated to measure all four bits, allowing us to resolve up to 15 photons, as shown in Fig. 4.3 (a).

Another approach performs a series of parity-like measurements, feeding forward the result of each bit measurement to determine the correct qubit rotation angle for the next measurement [Dassonneville et al., 2020]. However, any errors that occur during a single measurement are fed forward and corrupt any subsequent measurements, resulting in correlated errors. Our method, which independently measures each bit, is more suitable for error mitigation with post-processing techniques.

Each component of our measurement chain is susceptible to errors. The storage mode suffers from stochastic photon loss at a rate κ. The ancilla is vulnerable to dephasing and excitation decay during the mapping pulse $\mathrm{C}\hat{\mathcal{P}}_N$ as well as decay during the readout procedure. The probability of reading out the ancilla in the ground (excited) state when the readout should have yielded $e(g)$ is $\epsilon_{e(g)}$. The long lifetimes of the storage mode $T_1^s \approx 1\,\mathrm{ms}$ and ancilla $T_1^\sigma \approx 25\,\mu\mathrm{s}$ relative to the $2.9\,\mu\mathrm{s}$ duration of each measurement ensure that these errors are small. The dynamic reset protocol rarely leaves the ancilla in the excited state, and is not a significant source of error.

Counter-intuitively, a single error in the entire sequence can produce a result that differs from the correct value by up to eight photons. This happens because a single error can flip multiple bits in the photon number's binary decomposition. For example, the loss of a single photon from $|8\rangle = |1000\rangle$ after the measurement of $b_0 = 0, b_1 = 0, b_2 = 0$ results in $b_3(7) = 0$. Thus, the effect of such an error is to mistakenly read out $|8\rangle = |1000\rangle$ as $|0\rangle = |0000\rangle$. This error is illustrated by the green square in the right panel of Fig. 4.3 (b), which also shows the measurement outcomes when a single ancilla readout or storage decay error occurs.

4.5 Measurement Error Mitigation

Can we use our knowledge of the error mechanisms described above to improve the fidelity of our measurements? To answer this question, we use an approach known as error mitigation which relies on data post-processing to improve the quality of an ensemble of calculations or measurements. Error mitigation has been proposed [Temme et al., 2017,Li and Benjamin, 2017] and realized [Kandala et al., 2019] for use with variational algorithms. Here, our error mitigation efforts focus on an ensemble of measurements which sample from the population distribution of some quantum state.

In this section, we introduce an error model based on the error syndromes described in Section 4.4.3. This model is a function of several error rates, which must be calibrated. We discuss our calibration technique and compare the model to experimental results. Finally, we introduce the deconvolution techniques that use the calibrated model to mitigate the errors in the measurement and show an improvement on two sets of states.

4.5.1 Error model

Our goal is to characterize the errors so that we can model measurement outcomes. This is equivalent to finding the elements $\{F_i\}_{0\leq i\leq N_{\max}}$ of the positive operator-valued measure (POVM) that describes our measurement [Nielsen and Chuang, 2011, Maciejewski et al., 2020]. For instance, the bitwise measurement is composed of B individual bit measurements, each with a corresponding POVM $\{E^{(k)}_{b_k=0}, E^{(k)}_{b_k=1}\}$ where $E^{(k)}_{b_k} = \left(\hat{B}^{(k)}_{b_k}\right)^\dagger \hat{B}^{(k)}_{b_k}$. Each ideal F_i is then a product of POVM elements constituting the bitwise measurement:

$$F_{i=b_{B-1}\ldots b_0} = E^{(B-1)}_{b_{B-1}} \ldots E^{(0)}_{b_0}. \tag{4.10}$$

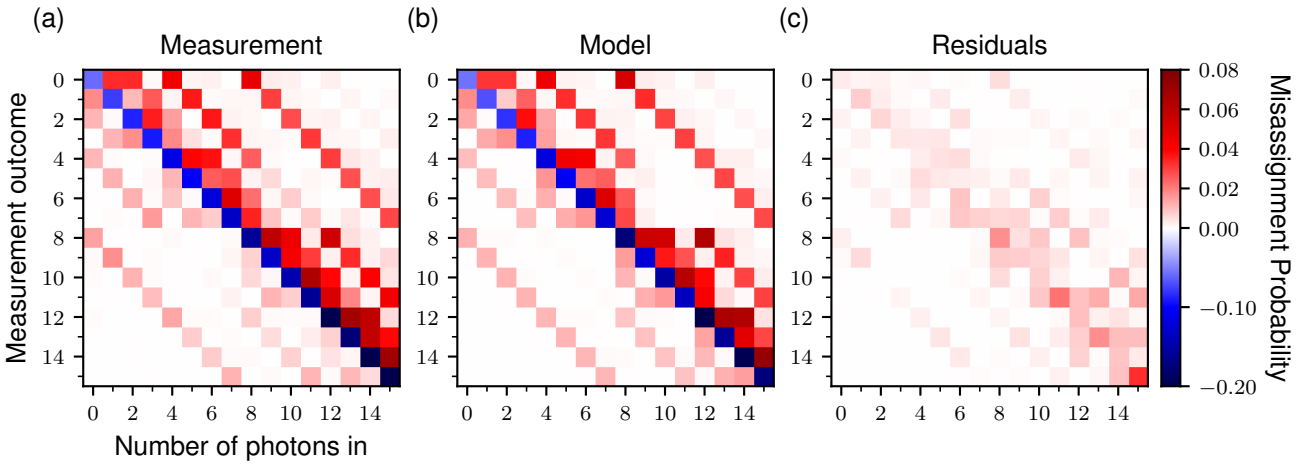


Figure 4.4: Measured error syndromes and hidden Markov model prediction. (a) Misassignment probabilities of prepared Fock states measured bitwise. (b) Misassignment probabilities predicted by the HMM. The HMM models the error syndromes depicted in Fig. 2 using the known storage mode loss rate κ and the calibrated error rates $\epsilon_g^{(k)}$, $\epsilon_e^{(k)}$. The HMM correctly predicts the structure of the errors (b) and agrees with the data down to the percent level (c). The negative values along the diagonal indicate the correct assignment probability deficit with each column summing to zero. The rows are the measured and modeled $\{F_i - \mathbf{1}_i\}$, respectively, where $\mathbf{1}_i$ is the matrix that is zero except for a one in its i^{th} diagonal entry. The residuals may arise from calibration errors or other error syndromes that are not modeled by the HMM. The residuals provide a typical bound of less than 1%, and no more than 4%, on the error syndromes not included in the HMM.

Note that the F_i and $E_{b_k}^{(k)}$ are diagonal because the bit measurement operators $\hat{B}_{b_k}^{(k)}$ are QND. Ideally, $F_i = |i\rangle\langle i|$, but our implementations of the $E_{b_k}^{(k)}$ have errors, so $(F_i)_{j,j}$ is the probability that the measurement detects i photons when there are j photons in the mode $P(\text{outcome} = i|\,|j\rangle)$.

Our system can be described with a hidden Markov model (HMM) that has been used for qubit readout [Dréau et al., 2013, Gammelmark et al., 2013, Ng and Tsang, 2014, Wölk et al., 2015, Martinez et al., 2020] and to improve the readout of qubits encoded in oscillators [Hann et al., 2018, Elder et al., 2020]. A HMM parametrizes the errors in the storage mode, known as transitions, and the fidelity of the measurements of each bit, known as emissions. The probability of a transition in the storage mode

during a time interval t from $|i\rangle$ to $|j\rangle$ is

$$T_{i,j}(\kappa t) = \binom{i}{j} \left(e^{\kappa t} - 1\right)^{i-j} e^{-i\kappa t}, \tag{4.11}$$

where κ is the decay rate of the mode [Hann et al., 2018]. The ancilla can decay during the controlled rotation $\mathrm{C}\hat{\mathcal{P}}^k$ and the readout procedure, leading to an incorrect bit measurement. Since the bit measurements are QND, only the diagonal entries $E^{(k)}_{b_k,i}$ are non-zero, even in the presence of errors. We modify the entries of these ideal POVM elements

$$E^{(k)}_{b_k,i} = \begin{cases} \left\lfloor \frac{i}{2^k} \right\rfloor \text{ is even} & \begin{cases} 1 - \epsilon_g^{(k)} & \text{for } b_k = 0 \\ \epsilon_g^{(k)} & \text{for } b_k = 1 \end{cases} \\ \left\lfloor \frac{i}{2^k} \right\rfloor \text{ is odd} & \begin{cases} \epsilon_e^{(k)} & \text{for } b_k = 0 \\ 1 - \epsilon_e^{(k)} & \text{for } b_k = 1 \end{cases} \end{cases} \tag{4.12}$$

to introduce error rates $\epsilon_{g(e)}$ describing errors in the $\mathrm{C}\hat{\mathcal{P}}^k$ mapping or ancilla readout.

We now integrate both error mechanisms into a single model describing the errors in the measurement. The bitwise measurement algorithm described in Section 4.4.3 alternates between entangling pulses and ancilla reset. The cavity is subject to spontaneous decay for the duration of the algorithm, punctuated by projective measurements of the entangled ancilla. This duration, however, is not deterministic due to the dynamic ancilla reset. Our HMM models the errors in the measurement by constructing an alternating chain of transition and emission (measurement) events. To calculate the $\{F_i\}$, we sum over all transition paths s_k between storage state $|j\rangle$ and measurement outcome i. Weighting these paths by the probability that the bit

measurements yield the binary decomposition of $i = b_3 b_2 b_1 b_0$ gives

$$(F_i)_{j,j} = \sum_{s_1 \ldots s_B} T_{j,s_1} \left(\kappa t^{(0)}\right) E^{(0)}_{b_0,s_1} T_{s_1,s_2} \left(\kappa t^{(1)} + b_0 \kappa t'\right) E^{(1)}_{b_1,s_2} \ldots T_{s_{B-1},s_B} \left(\kappa t^{(B-2)} + b_{B-1} \kappa t'\right) E^{(B-1)}_{b_{B-1},s_B}. \tag{4.13}$$

Each transition matrix T and POVM element $E_i^{(k)}$ are parametrized by the respective error rates $\kappa t^{(k)}$ and $\epsilon_g^{(k)}, \epsilon_e^{(k)}$, which are not necessarily the same for each bit. Note that the time between each bit measurement depends on each result, as a measurement of $b_k = 1$ requires an ancilla reset pulse followed by readout verification. We discuss how the dynamic ancilla reset statistics inform our choice of $\kappa t'$ in Appendix A.1. Eq. 4.13 provides the desired parametrized POVM elements, which, once calibrated, we use to mitigate errors in our measurement.

4.5.2 Model calibration and results

Now that we have a model for the errors, we must determine the values of the error parameters $\kappa t^{(k)}$ and $\epsilon_g^{(k)}, \epsilon_g^{(k)}$. The duration of each bit measurement is recorded shot-by-shot; the microwave pulses have fixed durations and the dynamic ancilla reset records the number of required attempts. Since measurements of the cavity's decay rate κ are straightforward, we only need to determine the $\epsilon_g^{(k)}, \epsilon_g^{(k)}$.

We calibrate the bit measurement errors by measuring a single bit on a basis of states. This characterizes the $\{E_i^{(k)}\}$ from which we extract the $\epsilon_g^{(k)}, \epsilon_g^{(k)}$. This procedure relies on the ability to prepare a basis of states with high fidelity, so that preparation errors do not pollute the measurement errors. We prepare Fock states by repeatedly converting two excitations in the ancilla into a single photon as detailed in [Gasparinetti et al., 2016, Elder et al., 2020]. We check the photon number with a series of selective pulses. If any fail to flip the ancilla, the storage mode is cooled to vacuum and the preparation protocol tries again. This method prepares high-fidelity Fock states, but does allow photon decay during the final check (see the

Table 4.1: HMM error parameters. The error σ of $\epsilon_g^{(k)}, \epsilon_e^{(k)}$ is calculated by propagating error in Eq. A.2. $1/\kappa \approx 1$ms is the storage cavity decay rate. The parenthetical additional time is added when the ancilla needs to be reset and is equal to $\kappa t'$, as discussed in Appendix A.1. The comma in the first two indicates $\kappa t^{(0)}$ for Fock and coherent states, respectively, which have different preparation methods.

Bit k	$\kappa t^{(k)}$ $(+\kappa t')$	$\epsilon_g^{(k)} \pm \sigma$	$\epsilon_e^{(k)} \pm \sigma$
0	0.0040, 0.0032	0.019 ± 0.002	0.029 ± 0.002
1	0.0034 (+0.0046)	0.014 ± 0.001	0.026 ± 0.001
2	0.0034 (+0.0046)	0.011 ± 0.001	0.035 ± 0.002
3	0.0034 (+0.0046)	0.013 ± 0.001	0.033 ± 0.002

Supplementary Material for [Elder et al., 2020]). Appendix A.2 contains additional details about the calibration process, including how we account for these preparation errors.

Once we have calibrated the error rates in the model in Table 4.1, we use Eq. 4.13 to calculate the $\{F_i\}$ and thus reconstruct the detector POVM. We compare the modeled POVM to the measured POVM in Fig. 4.4 by preparing the Fock states $\{|j\rangle\}_{0\leq j\leq 15}$ and measuring all four bits. The residuals are typically $\lesssim 1\%$, and never more than 4%, showing good agreement with the model and bounding unmodeled errors. We could in principle use the measured detector POVM to perform error mitigation, but we would still need to model transitions during the Fock state preparation and would not be able to confirm our understanding of the error mechanisms.

4.5.3 Calculation of information extracted by the detector

To be able to call our detector single-shot, it must extract a large fraction of the information it is designed to measure per shot. In particular, of the four bits it ideally measures, we would like to determine how many of these bits are lost to measurement errors. The difference of ideal and wasted bits gives the average number of bits extracted per measurement.

Our task is then to calculate the number of wasted bits in our measurement, which

is equivalent to number of extra bits needed to classify the input given a measurement result (or the number of bits remaining to be measured). This quantity is the entropy S_i [Shannon, 1948] of the distribution $P(|j\rangle\,|i) = P(\text{detector input } |j\rangle\,|\text{measurement result } i)$. We finally average over the results i to obtain $4 - \langle S_i \rangle$, the average number of bits extracted in a single shot of the measurement.

We begin by calculating the conditional probability

$$P(|j\rangle\,|i) = \frac{P(\text{result } i|\text{input } |j\rangle)P(|j\rangle)}{P(i)}. \tag{4.14}$$

Recalling that $P(\text{result } i|\text{input } |j\rangle) = C_{ij}$ and assuming a uniform prior $P(|j\rangle) = 1/N$ we simplify Eq. 4.14

$$\begin{aligned} P(|j\rangle\,|i) &= \frac{C_{ij}P(|j\rangle)}{\sum_k C_{ik}P(|k\rangle)} \\ &= \frac{C_{ij}}{\sum_k C_{ik}}. \end{aligned} \tag{4.15}$$

Finally, we write down the average entropy over all measurement outcomes, weighting by $P(\text{outcome } i)$

$$\begin{aligned} \langle S_i \rangle &= -\sum_{ij} P(|j\rangle\,|i) \log_2(P(|j\rangle\,|i))P(i) \\ &= -\frac{1}{N}\sum_{ij} C_{ij} \log_2\left(\frac{C_{ij}}{\sum_k C_{ik}}\right). \end{aligned} \tag{4.16}$$

Using the error parameters listed in Table 4.1 to construct C, we find $4 - \langle S_i \rangle_{C,\,\text{Fock}} = 3.14$ bits. We can also use the measurement results in Fig. 4.4 to determine $P(\text{result } i|\text{input } |j\rangle)$, yielding $4 - \langle S_i \rangle_{\text{Fock}} = 3.11$ bits. This number likely provides an upper bound of $\langle S_i \rangle$ as the check protocol for Fock states (described in Section 4.5.2) introduces a lengthy delay between state preparation and measurement.

As mentioned above, the preparation errors included in both of these calculations

likely inflate the number of bits the measurement itself wastes. We cannot deconvolve only the preparation error from the measurement results, but we can zero the contribution of preparation to $\kappa t^{(0)}$. We emphasize that $\kappa t^{(0)}$ will always contain a contribution from the previous experimental step, but we can construct a C confined to errors that solely occur during the measurement duration. This scenario yields $4 - \langle S_i \rangle_{\text{C, no prep}} = 3.17$ bits, which is likely a lower bound as the residuals in Fig. 4.4 show that C does not perfectly capture all errors.

Finally, we can use this method to estimate the single-shot error budget. With $\kappa t^{(k)}$ as in Table 4.1 and unit fidelity $\mathrm{C}\hat{\mathcal{P}}^k$ pulses and ancilla readout ($\epsilon_g^{(k)} = \epsilon_e^{(k)} = 0$) we find $4 - \langle S_i \rangle_{\text{only } \kappa t^{(k)}} = 3.72$ bits. The remaining information is lost to photon loss in the storage mode. This suggests that significant gains can be made by optimizing the $\mathrm{C}\hat{\mathcal{P}}^k$ pulses and ancilla readout before needing to decrease $\kappa t^{(k)}$.

The raw measurement fidelity of each state is at least 80% before error mitigation even for the most challenging input state $n = 15$. The measurement extracts on average between 3.11 and 3.17 of the four possible bits of information. This fraction is greater 3/4, suggesting that the bitwise measurement is single-shot.

4.5.4 Error deconvolution

Now that we have modeled the entire collection of $\{F_i\}$, we form a matrix $C_{ij} = (F_i)_{j,j}$, where $C - I$ is shown in Fig. 4.4(b), of conditional probabilities referred to as the "confusion matrix." This allows us to relate the measured populations $\mathbf{P}^{\text{meas}}$ to the state's ideal distribution $\mathbf{P}^{\text{ideal}}$ with a simple relation [Maciejewski et al., 2020]

$$\mathbf{P}^{\text{meas}} = C\mathbf{P}^{\text{ideal}}. \tag{4.17}$$

Once we have C, we invert it to solve the above equation for the mitigated probability vectors $\mathbf{P}^{\text{mit}} = C^{-1}\mathbf{P}^{\text{meas}}$. These vectors are properly normalized, but may have

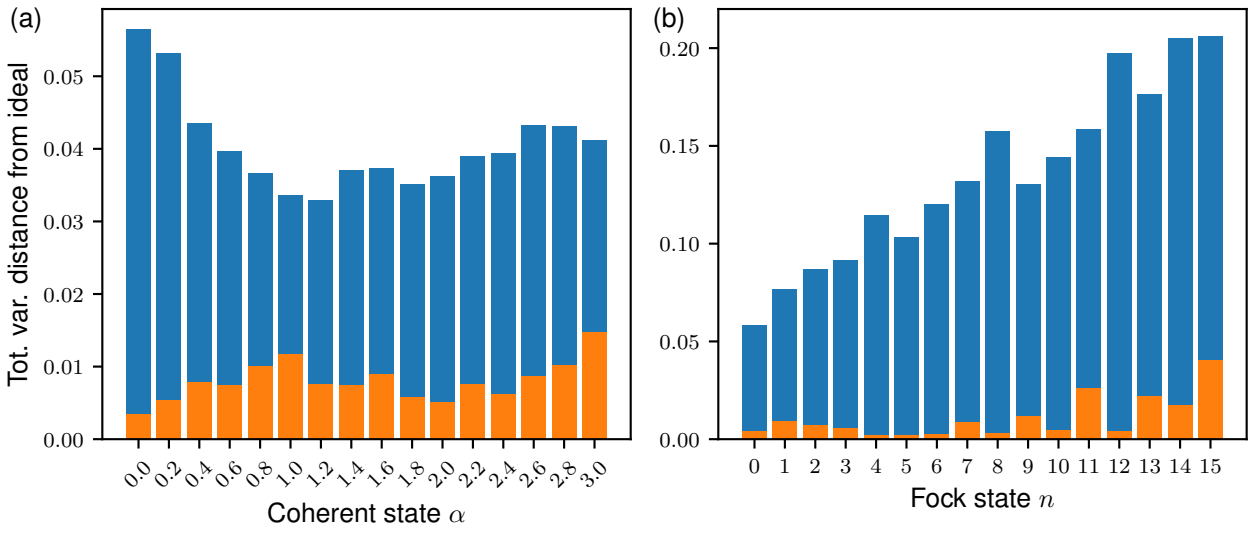


Figure 4.5: Error mitigation results for coherent (left panel) and Fock states (right panel). The α were chosen such that the Hilbert space can be truncated to the first 16 Fock states, with error less than 2.2%. For each prepared state $|\psi\rangle$, the error in the measurements $D_{TV}\left(\mathbf{P}^{\text{meas}}_{|\psi\rangle}, \mathbf{P}^{\text{ideal}}_{|\psi\rangle}\right)$ (blue) is compared to the error in the mitigated measurements $D_{TV}\left(\mathbf{P}^{\text{mit}}_{|\psi\rangle}, \mathbf{P}^{\text{ideal}}_{|\psi\rangle}\right)$ (orange). The large difference in overall error rates between Fock and coherent states results from the error mechanism to which each is most susceptible. Fock states transition to orthogonal states under photon loss, while coherent states are mainly affected by bit measurement errors. The single photon loss probability $\kappa t^{(k)} \approx 0.03$ is similar to the single bit error rates $\epsilon_e^{(k)} \approx .03, \epsilon_g^{(k)} \approx .01$.

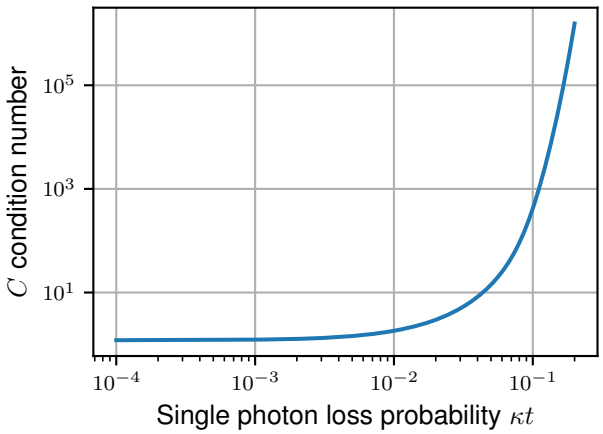


Figure 4.6: Condition number vs. storage mode error rate. As the cavity error rate increases, the condition number (ratio of largest to smallest singular value) of C diverges, indicating that the inversion of C is no longer numerically stable. For simplicity, we choose typical experimental values for the errors $\epsilon_g^{(k)} = 0.01$, $\epsilon_e^{(k)} = 0.03$ and sweep all $\kappa t^{(k)} = \kappa t'$ simultaneously.

negative entries. We correct the mitigation results by finding the closest vector $\mathbf{P}^{\text{mit}}$ with respect to the Euclidean norm that has non-negative entries [Maciejewski et al., 2020].

This protocol works in the range of error rates that allow robust inversion of C. In the limit of small error rates, C approaches the identity. As the error rates $\epsilon_g^{(k)}, \epsilon_e^{(k)}$, and $\kappa t^{(k)}$ become large, we expect that C will become ill-conditioned. For example, as the storage mode error rates $\kappa t^{(k)}$ approach 1, the probability C_{0n} of any photon state $|n\rangle$ decaying to $|0\rangle$ approaches 1. But since the column vectors of C are normalized to 1, they approach degeneracy and prevent the inversion of C. This exact scenario is illustrated in Fig. 4.6. Similarly, the condition number also diverges as $\epsilon_g^{(k)}, \epsilon_e^{(k)} \to 1$. We operate the small error rate regime in which C is robustly invertible.

The simple form of Eq. 4.17 offers the tantalizing possibility that we could avoid error modeling altogether by preparing a high-fidelity set of states, and directly inverting $\mathbf{P}^{\text{ideal}}$ to find C. Coherent states seem to be an ideal candidate due to the high fidelity with which we can prepare them, but suffer from ill-conditioned ideal populations $\mathbf{P}^{\text{ideal}}$, which make the determination of C highly sensitive to errors in

$\mathbf{P}^{\text{meas}}$ [Lundeen et al., 2009]. Furthermore, error modeling of the Fock states is required due to preparation error discussed in the previous section.

To quantify the quality of our measurement, we use the total variation distance metric to compare probability distributions $\mathbf{A}$ and $\mathbf{B}$,

$$D_{TV}(\mathbf{A}, \mathbf{B}) = \frac{1}{2} \sum_k |\mathbf{A}_k - \mathbf{B}_k| \,. \tag{4.18}$$

The factor of 1/2 accounts for the double-counting of probability differences.

To check the performance of the error mitigation and verify the error model, we apply it to the same Fock states we prepared in the calibration process, and to a set of coherent states with $0 \leq \alpha \leq 3$. The results are shown in Fig. 4.5. In both cases, the error mitigation includes photon losses occurring during state preparation (recall the selective pulse check discussed in Section 4.5.2). This is reasonable because the operation that directly precedes the bitwise measurement is known in advance as part of the pulse sequence.

The large difference in the pre-mitigated errors in Fig. 4.5 between the Fock and coherent states is due to the nature of these states. For the Fock states, a single lost photon transitions $|n\rangle$ to an orthogonal state $|m\rangle$, resulting in the maximum possible error. Furthermore, the transition rate increases with photon number, resulting in large pre-mitigation errors. On the other hand, coherent states are more robust to photon loss and do not transition to an orthogonal state, leaving bit measurement errors as the dominant error mechanism.

The error mitigation successfully reduces errors in all tested states to less than 5%, providing the most dramatic improvement to the Fock states, whose average error is reduced from 13.5% to 1.1%. The success of the error mitigation shows that we understand the error syndromes in our measurement and can effectively quantify them. This allows us to use the model to find the error rates required to meet a

desired measurement fidelity, which may vary from application to application.

4.5.5 Scalability of error mitigation

In the previous sections, we demonstrated an efficient measurement of the number of photons in a mode, and an effective method to mitigate the errors in the measurement. Our protocol can be performed simultaneously on a system with multiple storage modes, each dispersively coupled to their own ancilla qubit, to measure the joint photon number of the system. Imagine a system with M storage modes with uncorrelated measurement errors between the modes. Here, the entries of $\mathbf{P}^{\text{meas}}$ are elements of the joint probability distribution $P(n_1, n_2, \ldots, n_M)$. If we truncate the Hilbert space of each mode to some maximal number of photons N_{max}, then the length of $\mathbf{P}^{\text{meas}}$ is N_{max}^M. This exponentially large space suggests that our approach to error mitigation is not scalable to multiple modes.

However, it turns out that calculating any particular element of $\mathbf{P}^{\text{mit}}$ from $\mathbf{P}^{\text{meas}}$ is efficient, as we show below. Calculating particular elements of $\mathbf{P}^{\text{mit}}$ may be sufficient for problems in which the goal is to study features of an output distribution that can be altered by measurement errors. For instance, we may be interested in correcting the relative peak intensities of Franck-Condon factors [Wang et al., 2020, Clements et al., 2018], which only requires processing the measured peaks, not the entire Hilbert space. The resulting spectrum will have significant peaks at the same output photon numbers, with intensities adjusted to account for measurement errors, increasing the simulator's accuracy. This way, the simulator still does all the hard work of identifying the significant peaks; we only need to post-process the heights to mitigate measurement errors.

There are two properties that make calculating any element of $\mathbf{P}^{\text{mit}}$ efficient. The first is that the multi-mode confusion matrix is the Kronecker product of single-mode confusion matrices, meaning that it is efficient to calculate any element in the inverse

of the multi-mode confusion matrix. In general, though, C^{-1} contains an exponential number of nonzero matrix elements so writing down the entire matrix is not feasible. The second property is that the vector $\mathbf{P}^{\text{meas}}$ is necessarily sparse; assuming we do not perform an exponentially large number of measurements, most configurations $(n_1, n_2, \ldots, n_M)$ are never actually measured. Together, these two properties imply that any entry of $\mathbf{P}^{\text{mit}}$ can be efficiently computed,

$$\mathbf{P}_i^{\text{mit}} = \sum_{j \in \mathcal{S}} \left(C_1^{-1} \otimes C_2^{-1} \otimes \ldots \otimes C_M^{-1} \right)_{ij} \mathbf{P}_j^{\text{meas}} \tag{4.19}$$

where $\mathcal{S}$ is the set of nonzero entries in $\mathbf{P}^{\text{meas}}$, which is only polynomially large. We cannot use this fact to efficiently calculate the entire distribution $\mathbf{P}^{\text{mit}}$, however, because $\mathbf{P}^{\text{mit}}$ can contain exponentially many nonzero entries. In the case where simply adjusting peak intensities is insufficient, we have developed a method for expanding C^{-1} that mitigates errors up to a chosen order. This method is detailed in [Curtis et al., 2021].

4.6 Conclusion and outlook

Our bitwise measurement protocol utilizes the long lifetimes of 3D microwave cavities and the interactions generated by a Josephson junction to implement a high-fidelity, single-shot, microwave photon number-resolving detector. We use error mitigation techniques to improve the measurement fidelity of the Fock states by nearly an order of magnitude. Larger resolvable photon numbers are reachable by synthesizing additional $\mathrm{C}\hat{\mathcal{P}}^k$ pulses to measure more bits.

In addition to enabling error mitigation, our error model provides insight into the error budget of our PNR detector. For example, we estimate that with unit fidelity $\mathrm{C}\hat{\mathcal{P}}^k$ pulses and ancilla readout ($\epsilon_g^{(k)} = \epsilon_e^{(k)} = 0$) we can extract 3.72 bits of information in a single shot in our system with the current $\kappa t^{(k)}$. This suggests

that further optimization of the control pulses and ancilla readout can increase the information extracted per shot by up to 0.6 bits. Any additional improvements to the single-shot fidelity require decreasing $\kappa t^{(k)}$ by decreasing the pulse lengths or increasing cavity lifetimes. Adaptive check methods may also be able to improve single-shot fidelity at the expense of adding complexity to the error model.

Even with limited fidelity, optical PNR detectors have motivated the development of multiphoton quantum information processing protocols. Bringing high-fidelity PNR detectors to the microwave regime further bolsters the ability of cQED to perform photonic protocols, such as vibronic spectra simulations mentioned in Section 4.2. Additionally, this capability will motivate the development of algorithms that take advantage of the large bosonic Hilbert space available in cQED systems.

Chapter 5

Principles of error-detected gates

The bitwise measurement explored in Chapter 4 demonstrated that with careful calibration, we can deconvolve the effect of transmon and qubit errors in a measurement, on average. This technique is promising for experiments where the goal is to reconstruct a probability distribution that represents transition probabilities, but isn't generally applicable when we want to maximize the fidelity of a single measurement. For the rest of the dissertation, we shift focus to maximizing the fidelity of a single run of a quantum circuit via error detection.

Error detection interleaves measurements between gates whose result can tell us if something went wrong during the gate. This technique has even higher standards for the quality and non-invasiveness of our measurements. Failure to detect errors ruins gates immediately, causing potentially unrecoverable errors in logical qubits. It is also not sufficient for these measurements to be simply QND, which can still allow dephasing within the codespace, we need them to have a minimal, or deterministic, effect on our stored information.

Error detection differs from error correction in that the information stored in an error-detected circuit is lost when the error occurs. This means that a single error-detected circuit itself isn't useful for error correction, but a large number of error-

detecting circuits can be joined together in an error-correcting code to form a logical qubit [Wu et al., 2022, Teoh et al., 2023]. For a detailed and thorough introduction to these ideas, see James Teoh's excellent dissertation [Teoh, 2023].

This approach holds a large advantage over conventional approaches of joining large numbers of error-prone devices together because the extra information we extract from error detection makes decoding logical errors much easier. In this paradigm, error-correction based on higher-order codes (such as the surface code) of error-detected physical qubits has two error rates (thresholds) below which scaling the number of physical qubits decreases the overall logical error rate. The first of these rates, the erasure rate, is the probability that we detect an error in a physical qubit and erase it from the logical code by resetting it. This approach is especially powerful when uncorrectable leakage errors are detectable and can be flagged. The second is the remaining error rate in gates that aren't erased. In the case that leakage errors are efficiently converted into erasures, the remaining errors are Pauli errors in the code space. Both of these thresholds depend on the fraction of errors that can be detected [Wu et al., 2022]. This makes sense because as the fraction of detectable errors decreases to zero, we expect to recover traditional surface code thresholds. Notably, knowing *where* errors occur in the code allows up to $d-1$ erasures per cycle of error correction, an increase from the limit of $(d-1)/2$ Pauli errors in conventional surface codes [Teoh et al., 2023].

The error-detected physical qubits used in this dissertation are formed from a microwave cavity (or two) coupled to a transmon ancilla. We encode a logical qubit in the cavity mode(s) and use the transmon to decode and perform gates on the cavity. The transmon, however, has much faster decoherence rates and is the dominant contributor to infidelity in the resulting gate. We will mainly focus on detecting transmon errors in these chapters. Detection of errors in the logical qubit is a function of the code we use to encode the logical information. Error-detecting codes include dual

rail [Kubica et al., 2022, Teoh et al., 2023] and binomial and the generalized Fock $0N$ codes with code words $|0\rangle$ and $|N\rangle$ [Michael et al., 2016]. Generally, any code that satisfies the Knill-Laflamme condition is also an error-detecting code [Teoh, 2023].

In the next three chapters, we introduce error-detected gates and demonstrate a logical error-detected $R_Z^L(\theta)$ gate on a binomial qubit. From now on, we use superscript L to indicate rotations explicitly on logical qubits encoded in quantum error-correcting or error-detecting codes. Absence of superscript L indicates generality to logical or physical qubits. In this chapter, we introduce the idea of error-detected gates, show that their advantage results from our high-fidelity measurement capabilities. We also discuss the ingredients needed for an error-detected gate in the transmon-cavity architecture that we utilize to design our error-detected $R_Z^L(\theta)$ gate in Chapter 6. Chapter 7 extends our usage of the transmon as a qubit to a qutrit, which is an essential ingredient for our error-detected gate. In Chapter 8, we put to work our list of requirements and new qutrit design to and benchmark an error-detected $R_Z^L(\theta)$ rotation.

5.1 Sources of gate errors

Almost all quantum algorithms and error correction can be decomposed into a series of gates followed by a measurement. Let's take a look at a series of gates $\mathcal{U}_i$ acting sequentially on an ensemble of initial states

$$|\rho_{\text{initial}}\rangle \; \boxed{\mathcal{U}_1} \; \boxed{\mathcal{U}_2} \; \cdots \; \boxed{\mathcal{U}_n} \; |\rho_{\text{final}}\rangle$$

In the absence of any coupling between our quantum system and the environment, the gate propagators $\mathcal{U}_i$ are unitary. This means that the only source of error in the circuit is our ability to design the controls $\epsilon(t)$ that generate $\mathcal{U}_i\rho = U_i\rho U_i^\dagger$ to implement our desired unitary gates. The quality of our implementation is quantified by the

trace/unitary/control fidelity defined in 2.54. Maximizing the trace fidelity requires careful calibration of the Hamiltonian and response of the microwave components that produce and filter the drives.

Unfortunately, in reality, there is always unwanted coupling between our systems and the environment. We describe this coupling with jump operators J_i that can add noise to certain system parameters (such as resonant frequency noise, dephasing) or swap excitations between the system and environment (decay). The evolution in this case is not unitary, and is generated by the quantum Liouvillian in Eq. 3.15

$$\mathcal{U} = \mathcal{T} \exp\left(\int \mathcal{L}(t)dt\right). \tag{5.1}$$

On its own, this expression is not particularly helpful in revealing the effects of decoherence, but we can organize the Dyson series of the propagator as a sum of superoperators

$$\mathcal{U}(t,0)\rho = \sum_{p=0}^{\infty} \mathcal{G}_p(t,0)\rho \tag{5.2}$$

$$= \sum_{p=0}^{\infty} p_{\text{phys}}^p g_p(t,0)\rho \tag{5.3}$$

where $\mathcal{G}_p(t)$ is the time evolution when *exactly* p jumps of any type occur [Ma et al., 2020]. For small error rates, the probability of p jumps monotonically decreases, meaning that $\mathcal{G}_p(t)$ contribute less to the total evolution $\mathcal{U}(t,0)$ as p increases. In Eq. 5.3, we separate out a scale factor p_{phys}^p from $\mathcal{G}_p(t)$ that is the approximate probability that p jumps occur. Generally, $p_{\text{phys}}^p \propto (\kappa t)^p$ for some average decoherence rate κ.

Note that Eq. 5.2 is not necessarily the same as the Kraus representation of $\mathcal{U}(t,0)$, although some elements of the sum may have obvious Kraus representations. One

such term is $\mathcal{G}_0(t)$ which takes the form

$$\mathcal{G}_0(t,0)\rho = U_{\text{eff}}(t,0)\rho U_{\text{eff}}^\dagger(t,0) \tag{5.4}$$

where

$$U_{\text{eff}}(t,0) = \exp\left(-i\int_0^t H_{\text{eff}}(t_1)dt_1\right) \tag{5.5}$$

and the effective Hamiltonian $H_{\text{eff}}(t)$ in this case includes the non-Hermitian anti-commutator term in Eq. 3.15

$$H_{\text{eff}}(t) = H(t) - \frac{i}{2}\sum_l \kappa_l J_l^\dagger J_l \tag{5.6}$$

with jump operators J_l and decoherence rates κ_l. This "no-jump" term generates two effects: norm decay of pure states and no-jump backaction. The no-jump effects can be seen most clearly in the case of photon loss in a bosonic mode C. Consider a superposition of zero and n photons. The no-jump evolution generated by $\kappa_C C^\dagger C/2$ is

$$U_{\text{eff}}(t,0)\frac{1}{\sqrt{2}}\left(|0\rangle + |n\rangle\right) = \frac{1}{\sqrt{2}}e^{-\kappa_C t c^\dagger c/2}\left(|0\rangle + |n\rangle\right) \tag{5.7}$$

$$= \frac{1}{\sqrt{2}}\left(|0\rangle + e^{-\kappa_C tn/2}|n\rangle\right) \tag{5.8}$$

biases the state towards $|0\rangle$ and reduces the norm below 1 [Michael et al., 2016]. The new norm squared $\frac{1+2e^{-\kappa_C tn/2}+e^{-\kappa_C tn}}{2}$ represents the probability that no jumps occur in duration t.

To find the evolution under a fixed, non-zero number of jumps p, we average over

all possible trajectories by integrating over all jump times [Ma et al., 2020]

$$\mathcal{G}_p(t,0) = \int_0^t dt_p \cdots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 \mathcal{G}_0(t,t_p)\mathcal{S}\cdots\mathcal{S}\mathcal{G}_0(t_2,t_1)\mathcal{S}\mathcal{G}_0(t_1,0) \tag{5.9}$$

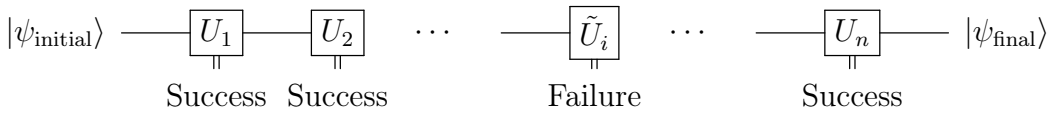
where $\mathcal{S}$ is the effect of discrete jumps from Eq. 3.15:

$$\mathcal{S}\rho = \sum_l \kappa_l J_l \rho J_l^\dagger. \tag{5.10}$$

The norm of each $\mathcal{G}_p(t,0)\rho$ represents the probability of p jumps occurring during the evolution of ρ.

This means that, unsurprisingly, all of the error in a gate $\mathcal{U}$ comes from the effects of discrete jumps and no-jump evolution. The goal of quantum error detection is then to filter out the contribution of $\mathcal{G}_p$ for $p \geq 1$ in NISQ-like applications or notify a decoder of an erroneous gate in real-time. If we know that no jumps have occurred in a duration dt, we can correct the no-jump evolution to $O(\kappa_C dt)$ with a unitary [Michael et al., 2016, Mirrahimi et al., 2014]. This is not done on average with error mitigation techniques like the ones discussed in Chapter 4, but by detecting jumps as they happen on a shot by shot basis. We do this by engineering $H_{\text{eff}}(t)$ such that jumps have a controlled and measurable effect on the system.

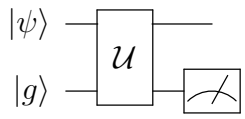
In other words, what we really want is some method to mark each gate in a string as successful or failed without disturbing the fidelity of states where all of these checks pass. When an attempted application of the channel $\mathcal{U}_i$ fails due to a jump, we should flag it



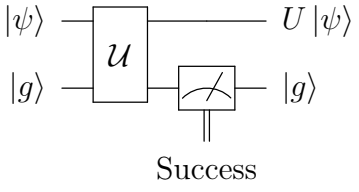
where the failed gate $\tilde{U}_i \rho \tilde{U}_i^\dagger$ represents a trajectory with at least one jump that contributes to the average evolution $\mathcal{U}_i\rho$. From now on, we use U to indicate a successful,

jump-free application $U|\psi\rangle = U_{\text{eff}}(t,0)|\psi\rangle / |\langle\psi|U_{\text{eff}}(t,0)^\dagger U_{\text{eff}}(t,0)\psi\rangle|$ of the channel $\mathcal{U}$ and $\tilde{U}$ to indicate a failed application with at least one jump. The exact form of $\tilde{U}$ depends on the timing of the jumps that occurred.

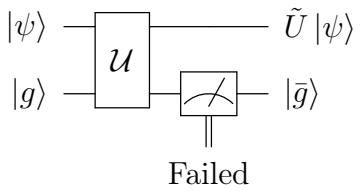
As stated in the introduction to this chaper, the standard cQED system is a cavity coupled to a transmon ancilla qubit, which enables gates and measurements on the linear cavity mode. When we perform a gate on a logical qubit encoded in the cavity mode, we often necessarily entangle the transmon with the logical information. This certainly happens when using optimal control, as discussed in Section 2.2.2. When optimizing these gates, we need to specify the evolution on both the cavity and transmon modes. Generally, we choose a set of state transfers such that the control pulses implement the identity on the transmon and the desired unitary on the cavity $|g\rangle \otimes |\psi\rangle \to |g\rangle \otimes U|\psi\rangle$. However, if the transmon decoheres during the gate, the evolution is no longer unitary and produces a random state in both the cavity and the qubit. But we know that a successful gate always takes $|g\rangle \to |g\rangle$, so adding a transmon measurement after the gate



will detect final states other than $|g\rangle$ that may herald errors that leave the transmon in a superposition. We want to figure out a way to design gates that maximizes the probability that a success results in a specific flag state, which we choose to be $|g\rangle$ for the rest of this dissertation. That is, the measurement projects $\mathcal{U}$ into a successful state $U|\psi\rangle$ when

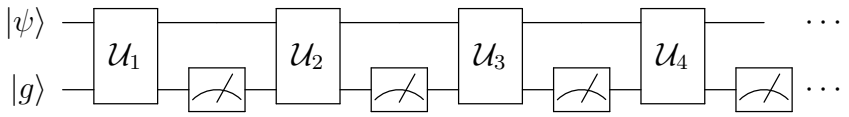


and failed when



where $|\bar{g}\rangle$ is a non-ground state of the transmon.

It is of utmost importance that we verify that any error detection we perform has minimal backaction on the information encoded in the cavity state $|\psi\rangle$. This way, when we concatenate gates and error detection (transmon measurements)



we don't add any infidelity beyond the unitary infidelity of the gates.

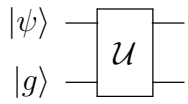
5.2 Error detection basics

Before designing error-detected gates, we want to check if the framework we described improves the fidelity of the gates we mark as successful. In particular, we want to answer a few questions:

- **What sets the rate at which failed gates slip through our error detection scheme?**
- **If we are relying on the transmon ancilla to flag gate errors, we need high-fidelity, QND transmon measurements. How accurate and QND are these measurements? What are typical values?**
- **What is the connection between gate success rate, detection fidelity, and accepted gate fidelity?**
- **How do undetected errors affect subsequent gates?**

5.2.1 The case where all errors are detectable

Let's start by finding a simple model for the state at the output of an error-detected gate. Suppose that we've designed our gate as discussed above so that successful gates leave the transmon in $|g\rangle$ and failed gates leave the transmon in other states. Then we can write the effect of $\mathcal{U}$ on a collection of initial states $\rho_0 = |g\rangle\langle g| \otimes |\psi\rangle\langle\psi|$



as

$$\mathcal{U}\rho_0 = P_{\text{success}}|g\rangle\langle g| \otimes U|\psi\rangle\langle\psi|U^\dagger + \mathcal{U}_{\text{fail}}\rho_0 \tag{5.11}$$

where P_{success} is the probability of no jumps occurring, producing a successful gate.

Now, we know that by design $\mathcal{U}_{\text{fail}}$ will map the transmon state $|g\rangle$ to some excited state, but how does $\mathcal{U}_{\text{fail}}$ act on the state $|\psi\rangle$? Since the jumps occur randomly, we assume that the net effect of all $\mathcal{G}_p(t,0)$ for $p \geq 1$ is to depolarize any input ρ_0 and produce a uniformly random output

$$\mathcal{U}_{\text{fail}}\rho_0 = \sum_{p=1}^{\infty} \mathcal{G}_{p\geq 1}(t,0)\rho_0 = \frac{p_{\text{phys}}}{d}\rho_{\bar{g}} \otimes \mathbf{I} + O(p_{\text{phys}}^2) \tag{5.12}$$

where $\mathbf{I}$ is the $d \times d$ identity matrix on the cavity space and $\rho_{\bar{g}}$ is a depolarized transmon state with $\langle g|\rho_{\bar{g}}|g\rangle = 0$ and $\text{Tr}(\rho_{\bar{g}}) = 1$. This assumption allows us to define a simple, trace-preserving channel for the gate $\mathcal{U}$ to first order in the physical failure rate

$$\mathcal{U}\rho_0 = (1 - p_{\text{phys}})|g\rangle\langle g| \otimes U|\psi\rangle\langle\psi|U^\dagger + \frac{p_{\text{phys}}}{d}\rho_{\bar{g}} \otimes \mathbf{I} \tag{5.13}$$

where a transmon measurement with result g projects into the successful state $U|\psi\rangle$

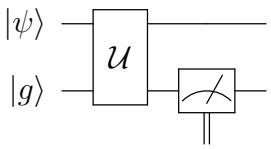
with probability

$$P_{\text{success}} = \text{Tr}\left((|g\rangle\langle g| \otimes \mathbf{I})\mathcal{U}\rho_0\right) = (1 - p_{\text{phys}}). \tag{5.14}$$

However, as we learned in Chapter 4, measurements aren't perfect, and the transmon measurement operators M_k rarely project completely into state $|k\rangle$. Furthermore, the M_k may even have residual non-QNDness in the form of off-diagonal elements that induce transitions between transmon states. For now, let's consider the diagonal measurement operator

$$M_g = \sqrt{P(g|\,|g\rangle)}|g\rangle\langle g| + \sqrt{P(g|\,|\bar{g}\rangle)}\sum_{k\neq g}|k\rangle\langle k| \tag{5.15}$$

that correctly detects the ground state with probability $P(g|\,|g\rangle)$ but also mistakes non-ground states $|\bar{g}\rangle$ as ground with probability $P(g|\,|\bar{g}\rangle)$. With this measurement operator, we can fully write the output ρ_{success} of a gate flagged successful



and see the effect of both gate failure rate and M_g. The average output state

$$\rho'_{\text{success}} = M_g\mathcal{U}\rho_0 M_g = (1 - p_{\text{phys}})P(g|\,|g\rangle)|g\rangle\langle g| \otimes U|\psi\rangle\langle\psi|U^\dagger + \frac{p_{\text{phys}}P(g|\,|\bar{g}\rangle)}{d}\rho_{\bar{g}} \otimes \mathbf{I} \tag{5.16}$$

(dropping the $O(p_{\text{phys}}^2)$ correction from now on) is unnormalized because projective measurements are not trace preserving. Before we normalize, notice that the rightmost term's coefficient is the *product* of two presumably small rates p_{phys} and $P(g|\,|\bar{g}\rangle)$. As a result, two errors must occur for a failed gate to occur in our circuit: the gate

failing itself and a measurement error. See Fig. 5.1 for a visualization of the error pathways in an error-detected gate. After renormalizing, we have

$$\rho_{\text{success}} = (1-\epsilon_{\text{pass}})|g\rangle\langle g| \otimes U|\psi\rangle\langle\psi|U^\dagger + \frac{\epsilon_{\text{pass}}}{d}\rho_{\bar{g}} \otimes \mathbf{I} \tag{5.17}$$

where ϵ_{pass} is the rate at which failed gates pass our check

$$\epsilon_{\text{pass}} = \frac{p_{\text{phys}}P(g|\,|\bar{g}\rangle)}{(1-p_{\text{phys}})P(g|\,|g\rangle) + p_{\text{phys}}P(g|\,|\bar{g}\rangle)} \sim \frac{p_{\text{phys}}P(g|\,|\bar{g}\rangle)}{P(g|\,|g\rangle)}. \tag{5.18}$$

The normalization denominator of ϵ_{pass} is also the probability that we accept a gate

$$P_{\text{success}} = (1-p_{\text{phys}})P(g|\,|g\rangle) + p_{\text{phys}}P(g|\,|\bar{g}\rangle) \tag{5.19}$$

as successful. We call this the "success rate" and p_{phys} the "failure rate."

5.2.2 The case where only one jump is detectable

However, we will see that it is often not possible to design a gate such that *all* trajectories with more than zero jumps are detectable. Let's consider the more reasonable case where we can only design a gate $\mathcal{U}$ where one jump is detectable. For example, the aforementioned binomial code allows error detection via a parity measurement after one jump that switches its parity to odd. A second jump switches the parity back to even causing a parity measurement to flag no error. Consider a $\mathcal{U}$ constructed such that only trajectories with exactly one jump leave the transmon in $|\bar{g}\rangle$, but for $p > 1$ jumps depolarize the transmon as well. That is

$$\mathcal{G}_1(t,0)\rho_0 = \frac{p_{\text{phys}}}{d}\rho_{\bar{g}} \otimes \mathbf{I} \tag{5.20}$$

$$\sum_{p=1}^{\infty}\mathcal{G}_{p\geq 2}(t,0)\rho_0 = \frac{p_{\text{phys}}^2}{dq}\mathbf{I}\otimes\mathbf{I} + O(p_{\text{phys}}^3) \tag{5.21}$$

to second order in p_{phys} with q the size of the transmon Hilbert space. This means that there is a non-zero probability that failed gates with two or more jumps will be flagged as successful, even without any measurement errors

$$\mathcal{U}_{\text{fail}}\rho_0 = \frac{p_{\text{phys}}}{d}\rho_{\bar{g}} \otimes \mathbf{I} + \frac{p_{\text{phys}}^2}{dq}\mathbf{I} \otimes \mathbf{I} + O(p_{\text{phys}}^3). \tag{5.22}$$

See Fig. 5.2 for a visualization of the failure pathways when two or more jumps cannot be reliably detected.

We see clearly that the states with two jumps can leak through detection by following the same procedure as before to find the un-normalized final state

$$\rho'_{\text{success}} = P(g|\,|g\rangle)|g\rangle\langle g| \otimes \left((1 - p_{\text{phys}} - p_{\text{phys}}^2)U|\psi\rangle\langle\psi|U^\dagger + \frac{p_{\text{phys}}^2}{dq}\mathbf{I}\right) + P(g|\,|\bar{g}\rangle)\frac{p_{\text{phys}}}{d}\rho_{\bar{g}} \otimes \mathbf{I}. \tag{5.23}$$

to second order in the small error rates $P(g|\,|\bar{g}\rangle)$ and p_{phys}. The normalization factor is equal to the success rate

$$\text{Tr}(\rho'_{\text{success}}) = P(g|\,|g\rangle)\left(1 - p_{\text{phys}} - p_{\text{phys}}^2 + \frac{p_{\text{phys}}^2}{q}\right) + P(g|\,|\bar{g}\rangle)p_{\text{phys}}. \tag{5.24}$$

Normalizing $\rho_{\text{success}} = \rho'_{\text{success}}/\,\text{Tr}(\rho'_{\text{success}})$ doesn't produce as tidy of an expression as it does in Eq. 5.17, but even unnormalized, we see that as p_{phys} increases, at some point $p_{\text{phys}}^2 > p_{\text{phys}}P(g|\,|\bar{g}\rangle)$ and two jump errors become the dominant source of infidelity over misclassified single jump errors.

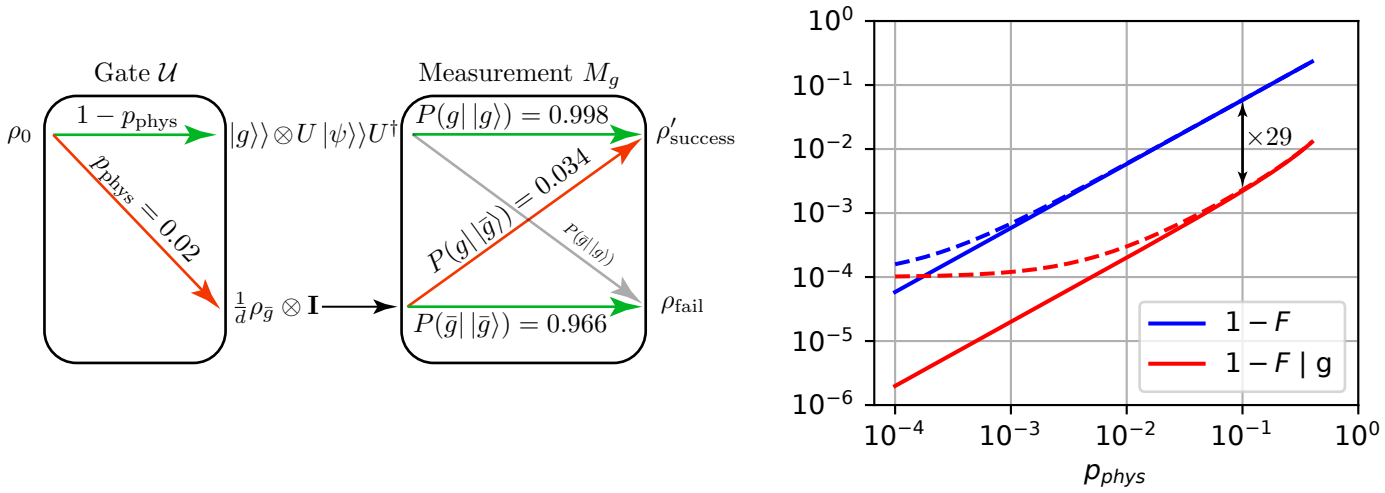


Figure 5.1: Pathways and fidelity of error detection when all jumps are detectable. **Left**: error pathways in both the application of the gate and the error detection. The probability of each pathway is labeled with a corresponding output state. There is a small probability $P(\bar{g}|\,|g\rangle) = 0.002$ that successful gates are discarded as failed gates. **Right**: infidelities $1 - \bar{F}_g(U_t, \mathcal{U})$ with (red) and without (blue) error detection for two cases: $\bar{F}_g(U_t, U) = 1$ (solid line) and $\bar{F}_g(U_t, U) = 0.9999$ (dashed line). The factor of improvement $p_{\text{phys}}/\epsilon_{\text{pass}} = 29$ is constant for the unity fidelity case, and is maximized for the non-unity case at higher p_{phys}. The quantities used to generate these lines are the ones labeled on the left and are found in Table 5.1. The $1 - \bar{F}_g(U_t, \mathcal{U})$ are plotted in the regime where the approximation $p_{\text{phys}} \ll 1$ is valid.

5.3 Fidelity of error-detected gates

5.3.1 Fidelity when all jumps are detectable

While the process we described above to find ρ_{success} is not trace-preserving and thus not a valid quantum channel, we can still compute the overall fidelity of the accepted gates by tracing out the qubit. Let's start with the fidelity of the effective channel that produced the final state in Eq. 5.17, which detects all numbers of jumps

$$\mathcal{U}_{\text{eff}}\rho = (1-\epsilon_{\text{pass}})U\rho U^\dagger + \frac{\epsilon_{\text{pass}}}{d}\mathbf{I}. \tag{5.25}$$

Unlike the channel $\mathcal{U}$, the effective channel $\mathcal{U}_{\text{eff}}$ is only defined on the cavity because we are measuring out the transmon. We use Eq. 2.58 to compute the average gate fidelity of $\mathcal{U}_{\text{eff}}$ to a target cavity unitary U_t

$$\bar{F}_g(U_t, \mathcal{U}_{\text{eff}}) = (1-\epsilon_{\text{pass}})\bar{F}_g(U_t, U) + \frac{\epsilon_{\text{pass}}\,(1+d^2)}{d^2(d+1)}, \tag{5.26}$$

where we have used the fact that $\sum_j B_j = d$. This expression allows us to compute the error-detected gate fidelity from the control fidelity of the underlying unitary and the gate and measurement failure rate ϵ_{pass}. The fidelity scales linearly in p_{phys}, as shown in Fig. 5.1 for several different control fidelities. Without error detection, the fidelity of channel 5.13 has the exact same form as Eq. 5.26, but scales linearly with p_{phys} instead of ϵ_{pass}

$$\bar{F}_g(U_t, \mathcal{U}) = (1-p_{\text{phys}})\bar{F}_g(U_t, U) + \frac{p_{\text{phys}}\,(1+d^2)}{d^2(d+1)}. \tag{5.27}$$

The maximum infidelity reduction occurs at unity control fidelity $\bar{F}_g(U_t, U) = 1$ by a factor of $p_{\text{phys}}/\epsilon_{\text{pass}} \sim P(g|\,|g\rangle)/P(g|\,|\bar{g}\rangle)$. This ratio of true to false gate success sets the improvement we can expect from error detection. In the next section, we charac-

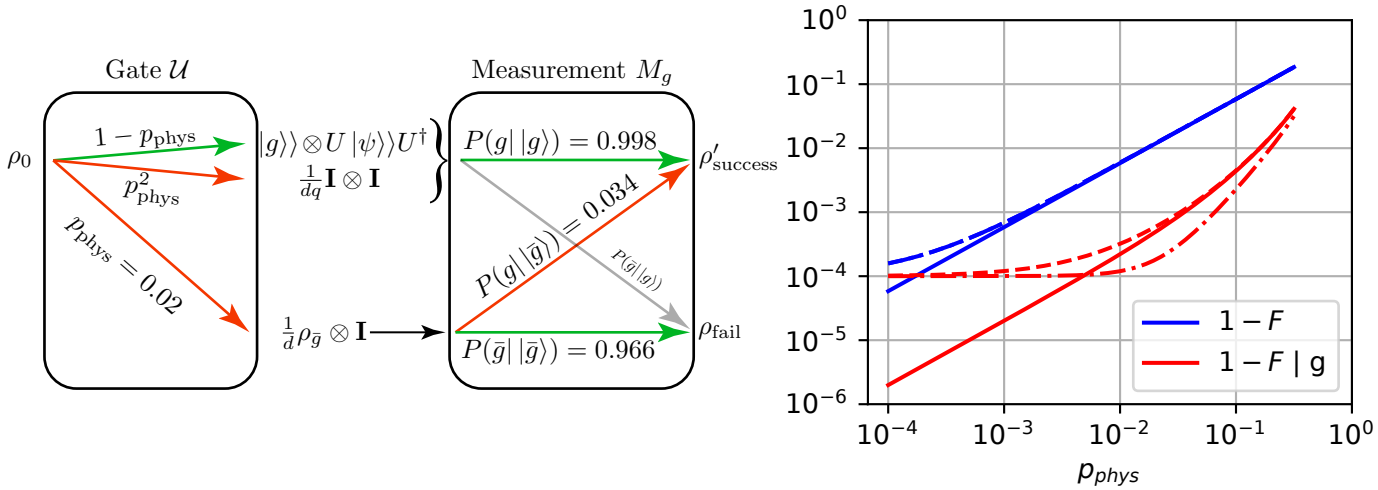


Figure 5.2: Pathways and fidelity of error detection when more than one jump is undetectable. This figure has the same format as Fig. 5.1. On the **left**, we have added an undetectable error pathway with rate p_{phys}^2. On the **right**, we see that the dashed line for the error-detected infidelity with $\bar{F}_g(U_t, U) = 0.9999$ does not achieve the same factor of improvement as in Fig. 5.1 due to the presence of undetected errors. These undetected errors have a non-linear effect on the infidelity, resulting in a slope greater than one. This is most clearly demonstrated by the dot-dashed line where $P(g|\,|\bar{g}\rangle) = 0$ to clearly show that p_{phys}^2 becomes the dominant source of error for larger p_{phys}. Again the plot is truncated to maintain the accuracy of our approximations.

terize both of these rates, finding that they are small and the ratio $P(g|\,|g\rangle)/P(g|\,|\bar{g}\rangle)$ approaches 30.

5.3.2 Fidelity when only one jump is detectable

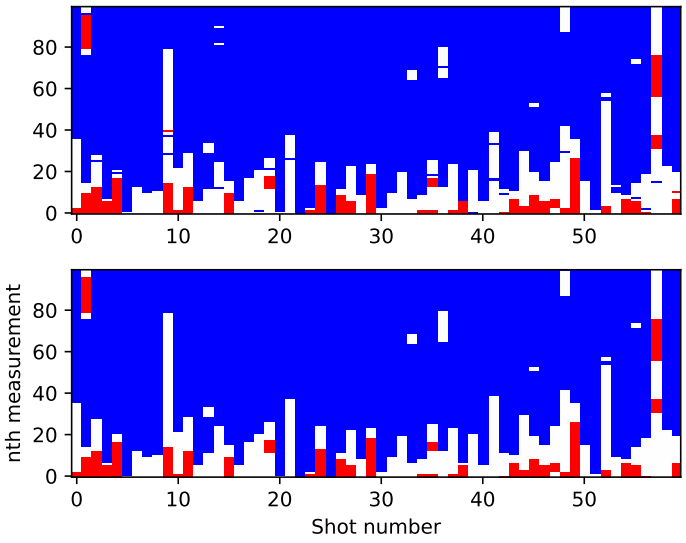
Repeating the same procedure for the more realistic case where only one jump is detectable we find the ϵ_{pass} in $\mathcal{U}_{\text{eff}}$ to be

$$\epsilon_{\text{pass}} = \frac{p_{\text{phys}}^2 P(g|\,|g\rangle)/q + p_{\text{phys}} P(g|\,|\bar{g}\rangle)}{P(g|\,|g\rangle)(1 - p_{\text{phys}} - p_{\text{phys}}^2) + p_{\text{phys}}^2 P(g|\,|g\rangle)/q + p_{\text{phys}} P(g|\,|\bar{g}\rangle)} \tag{5.28}$$

$$\sim \frac{p_{\text{phys}}^2 P(g|\,|g\rangle)/q + p_{\text{phys}} P(g|\,|\bar{g}\rangle)}{P(g|\,|g\rangle)}. \tag{5.29}$$

From Eq. 5.26, it's clear that for $p_{\text{phys}} \ll P(g|\,|\bar{g}\rangle)$ the failure misclassification dominates the infidelity, but the error from multiple undetected jumps p_{phys}^2 overtakes it as $p_{\text{phys}} \sim P(g|\,|\bar{g}\rangle)$. We see this trend borne out in Fig. 5.2. $O(p_{\text{phys}}^2)$ infidelity scaling

is then good evidence that erroneous trajectories in $\mathcal{G}_1(t,0)$ are being filtered out of ρ_{success} and is a signature we will look for in later sections.



Probability	Value
$P(g\|\,\|g\rangle)$	0.998
$P(\bar{g}\|\,\|g\rangle)$	0.002
$P(g\|\,\|\bar{g}\rangle)$	0.034
$P(\bar{g}\|\,\|\bar{g}\rangle)$	0.966
$P(\|g\rangle \to \|g\rangle)$	0.999
$P(\|g\rangle \to \|e\rangle)$	0.001
$P(\|e\rangle \to \|g\rangle)$	0.090
$P(\|e\rangle \to \|f\rangle)$	0.017
$P(\|f\rangle \to \|e\rangle)$	0.133

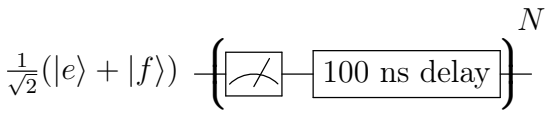
Figure 5.3 & Table 5.1: Transmon assignment fidelities and QNDness. **Top**: 60 series of measurements where the transmon was prepared in $\frac{1}{2}(|e\rangle + |f\rangle)$ and then measured 100 times. Blue indicates a measurement result corresponding to $|g\rangle$, white to $|e\rangle$, and red to $|f\rangle$. One can see long series of consistent measurements interrupted by different outcomes due to measurement errors. The **bottom** figure is a prediction of the underlying state producing each measurement result in the top figure. This prediction was produced by training a HMM on the measurement outcomes to find the underlying states, and the assignment and transition probabilities listed in the table. Errors are not provided, but could be found with resampling techniques.

5.4 QND transmon measurements

For the scheme described here, we need $P(g|\,|g\rangle) \sim 1$ and $P(g|\,|\bar{g}\rangle) \ll 1$. We first want to verify that our transmon measurements are QND on both the transmon and cavity so that the output of the error-detected gate isn't affected by the measurement. Without both of these conditions, we cannot concatenate gates and error detection is dead on arrival. Transmon QNDness is important so that the measurement doesn't heat the successful flag state $|g\rangle$ to $|e\rangle$ or $|f\rangle$, spoiling the next gate in a sequence.

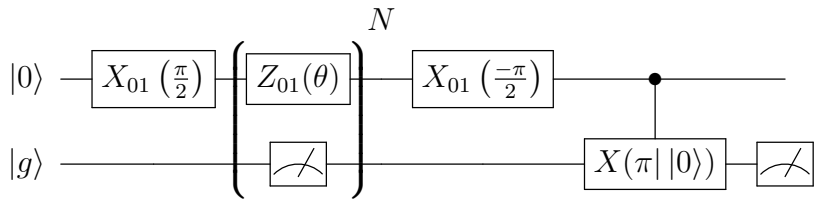
To find $P(g|\,|g\rangle)$ and $P(g|\,|\bar{g}\rangle)$, we perform a simple measurement where we pre-

pare the transmon and then measure it repeatedly. This provides a trajectory of states with which we can train a hidden Markov model [Gammelmark et al., 2014, Hann et al., 2018, Martinez et al., 2020] to find the transition rates (non-QNDness) and measurement fidelities $P(g|\,|g\rangle) \sim 1$ and $P(g|\,|\bar{g}\rangle) \ll 1$. See Appendix B for a copy of the code used. The experiment circuit is simple



repeatedly measuring the initial superposition N times to track jumps and measurement misassignments. The assignment and transition probabilities are listed in Table 5.1. The ratio $P(g|\,|g\rangle)/P(g|\,|\bar{g}\rangle) = 29$ suggests that error detection will improve the fidelity of accepted gates up to a factor of nearly 30. It's worth noting that these values allow us to specify the measurement effects $E_g = M_g^\dagger M_g$, which does not uniquely determine M_g. However, for simplicity we often assume M_g is QND, and therefore diagonal.

To check how QND the transmon measurement is on states stored in the cavity mode, we perform a cavity Ramsey measurement similar to the one in Section 3.1



replacing the idle unitary evolution with N repeated transmon readouts whose results we discard. The residual cross-Kerr between the populated readout resonator and the cavity produces a small frame rotation $Z(\theta)$ of $\theta = 0.013 \times 2\pi$ radians. See Fig. 5.4 for data and fit.

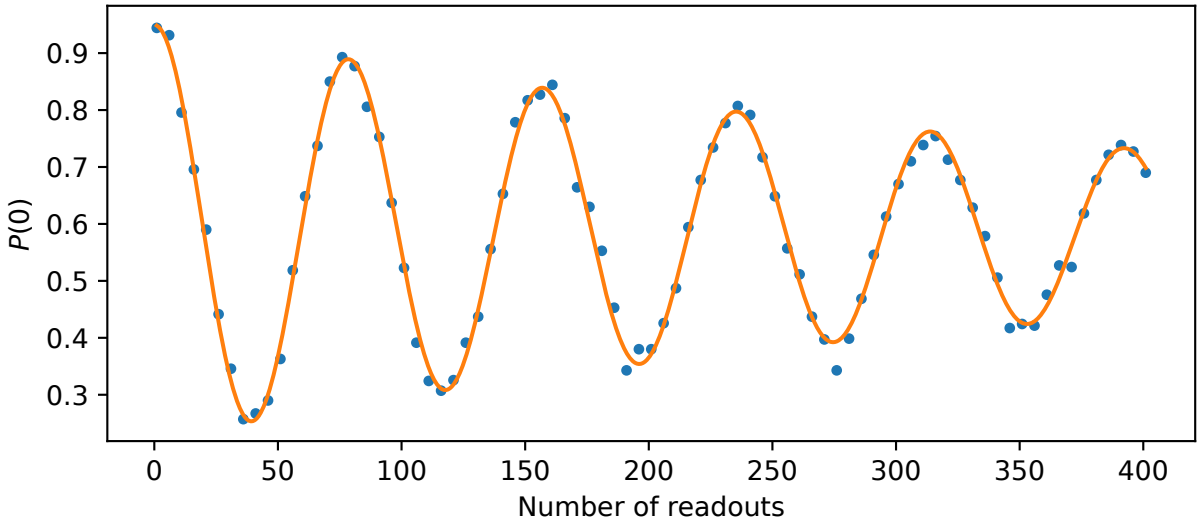


Figure 5.4: Fock readout Ramsey. Instead of a delay between $\pi/2$ pulses, we insert readouts. Each readout repetition is 2900 ns. From the fit, we find $T_2 = 1.26$ ms and a deterministic phase rotation per readout of $0.013 \times 2\pi$ radians. The T_2 here is comparable to that of Fig. 3.3 ($T_2 = 1.1$ ms) showing that the transmon measurement is largely QND with respect to the cavity.

5.5 Recap

- **What sets the rate at which failed gates slip through our error detection scheme?** In the gate design assumed above, $\mathcal{U}_{\text{fail}}$ leaves the transmon ancilla in one of the excited states $\rho_{\bar{g}}$. The rate at which we detect these errors is then the measurement fidelity of distinguishing the states $|g\rangle$ and $|\bar{g}\rangle$. The probability of a failed gate appearing in an ensemble of accepted gates is then the probability of both gate failure and measurement misassignment occurring. We call this rate

$$\epsilon_{\text{pass}} \sim \frac{p_{\text{phys}} P(g|\,|\bar{g}\rangle)}{P(g|\,|g\rangle)}. \tag{5.30}$$

In most cases, we design our gates to detect only one jump, leaving us blind to multiple jumps. The "pass" rate in this case is

$$\epsilon_{\text{pass}} \sim \frac{p_{\text{phys}}^2 P(g|\,|g\rangle)/q + p_{\text{phys}} P(g|\,|\bar{g}\rangle)}{P(g|\,|g\rangle)}. \tag{5.31}$$

- **If we are relying on the transmon ancilla to flag gate errors, we need high-fidelity, QND transmon measurements. How accurate and QND are these measurements? What are typical values?** The measured fidelity in our experiment of correctly labeling $|\bar{g}\rangle$ as a failed gate is $P(\bar{g}|\,|\bar{g}\rangle) = 0.966$. The rate at which we mistake a failed state for a successful one is $P(g|\,|\bar{g}\rangle) = 0.034$. These measurements are very QND on the transmon, exciting it out of its ground state with $P(|g\rangle \rightarrow |e\rangle) = 0.001$. Transmon measurements do not substantially affect the T_2 of the cavity mode, as shown in Fig. 5.4. They do produce a deterministic frame rotation that we can track by updating the phase of our local oscillators.

- **What is the connection between gate success rate, detection fidelity, and accepted gate fidelity?** Gates fail when jumps occur in the cavity or transmon mode. These jumps occur at rate p_{phys}, but can be detected. The rate at which we accept gates, the success rate, depends on how many jumps we can detect, but is generally the sum of the probability that no jumps occur (order unity) and undetectable jumps occur (small). Pathways for gates to be labeled successful are shown in Figs. 5.1 and 5.2. The fidelity of a single error-detected gate depends on ϵ and the control fidelity

$$\bar{F}_g(U_t, \mathcal{U}_{\text{eff}}) = (1-\epsilon)\bar{F}_g(U_t, U) + \frac{\epsilon\,(1+d^2)}{d^2(d+1)}. \tag{5.32}$$

where $\epsilon = \epsilon_{\text{pass}}$ when we utilize error detection and $\epsilon = p_{\text{phys}}$ when we don't.

- **How do undetected errors affect subsequent gates?** There are two types of undetected errors: misclassified gates and induced errors where the transmon readout following a successful gate excites the transmon out of $|g\rangle$. Failed gates that are misclassified as successful ones depolarize the cavity state and are not detectable during subsequent gates. These types of errors ruin a circuit, causing a logical error if a qubit is encoded in the cavity. If a transmon readout flags a gate as successful, but induces a transition to an excited state, the error can usually be detected after the following gate (this often depends on the specific gate implementation). When utilizing error detection within a larger logical qubit, the decoder can benefit from delayed error detection [Chang et al., 2023].

Chapter 6

Design of error-detected gates

Now that we are convinced that error-detected gates are useful and can provide a large increase in fidelity, we need to design a gate that detects jumps. To start, we will attempt to design gates that detect up to one jump; with current decoherence and gate times, the probabilities of double jumps are small compared to single jumps and misassignment [Teoh et al., 2023]. The three main sources of decoherence whose induced jumps we want to detect are

1. Cavity photon loss $\mathcal{D}[\sqrt{\kappa}C]$, with typical $1/\kappa \sim 1$ ms. We leave detection of cavity photon loss up to the code used to encode a qubit in the cavity. In the experiments that follow, we use the binomial/kitten code with code words $|0_L\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |4\rangle), |1_L\rangle = |2\rangle$. This is an error-correcting and detecting code up to losing one photon. Single photon loss switches parity to odd, which we can measure using the techniques of Chapters 3 and 4. We will not explicitly discuss cavity error detection in this chapter, although it will make a surprise appearance in Chapter 8.

2. Transmon excitation decay $\mathcal{D}[\sqrt{\kappa}A]$. Due to the transmon's anharmonicity, the decay rates between adjacent states do not scale linearly with excitation number as in a linear cavity. We often write two Lindblad jump operators for the

lowest two excited states $\mathcal{D}[\sqrt{\kappa_{ge}}|g\rangle\langle e|]$, $\mathcal{D}[\sqrt{\kappa_{ef}}|e\rangle\langle f|]$ because each transition couples of a different bath of photons. For the hardware used in this chapter, we typically find $1/\kappa_{ge} \sim 60$ μs and $1/\kappa_{ef} \sim 40$ μs. There is not an explicit two photon coupling to the environment, so multi-photon dissipation such as $\mathcal{D}[\sqrt{\kappa_{gf}}|g\rangle\langle f|]$ can be neglected for transmons [Koch et al., 2007].

3. Transmon dephasing $\mathcal{D}[\sqrt{2\Gamma^{\phi}}A^{\dagger}A]$. Again, we often write this as a sum of dephasing in two subspaces $\mathcal{D}\left[\sqrt{2\Gamma_e^{\phi}}|e\rangle\langle e|\right] + \mathcal{D}\left[\sqrt{2\Gamma_f^{\phi}}|f\rangle\langle f|\right]$. Note here that we absorb the matrix element of $A^{\dagger}A$ into Γ_n^{ϕ}. Typically, $1/\Gamma_e^{\phi}, 1/\Gamma_f^{\phi} > 100$ μs. The pure dephasing time $1/\Gamma_n^{\phi}$ is determined via $T_1^{(n)}$ measurements of $|n\rangle$ and $T_2^{(gn)}$ measurements of coherence times of superpositions $|g\rangle + |n\rangle$. Then use $\Gamma_n^{\phi} = 1/T_2^{(gn)} - 1/(2T_1^{(n)})$ to find the pure dephasing rate. Of all the decoherence mechanisms, it will require the most effort to detect deleterious dephasing jumps.

6.1 How to detect transmon decay?

Since we are interested in only detecting single jumps, our task is to design the $\mathcal{G}_1(t,0)$ with jump operator $\sqrt{\kappa}A$ such that gates with transmon excitation decay

$$\mathcal{G}_1(t,0)|g\rangle\langle g| \otimes |\psi\rangle\langle\psi| = \rho_{\bar{g}} \otimes \rho_{\text{fail}} \tag{6.1}$$

fail with the transmon in an excited state. However, we typically implement our gates with a Hamiltonian that drives the transmon's $g-e$ transition such as in Eqs. 2.36 and 2.47. This means that $J = \sqrt{\kappa_{ge}}|g\rangle\langle e|$ jumps interrupt the no-jump evolution of

the gate at random times

$$\mathcal{G}_1(t,0)\rho = \int_0^t dt_1 \mathcal{G}_0(t,t_1)\mathcal{S}\mathcal{G}_0(t_1,0)\rho \tag{6.2}$$

$$= \kappa_{ge}\int_0^t dt_1 U_{\text{eff}}(t,t_1)|g\rangle\langle e|U_{\text{eff}}(t_1,0)\rho U_{\text{eff}}(t_1,0)^\dagger|e\rangle\langle g|U_{\text{eff}}(t,t_1)^\dagger, \tag{6.3}$$

using the definitions from Section 5.1 and where we have assumed, for the moment, that the only error mechanism is transmon decay. From this expression, it's clear that while jumps interrupt the evolution of U_{eff}, the drives still continue to produce additional evolution because the jump $|g\rangle\langle e|$ leaves the transmon in the $g-e$ computational subspace. Since we cannot know when the jump occurred, we cannot engineer controls such that the transmon's final state is always $|\bar{g}\rangle$. Partial error detection is still possible though because any measurement result of $|\bar{g}\rangle$ indicates an error, but we cannot trust that a measurement of $|g\rangle$ yields a successful gate due to a jump leaving the transmon in a superposition state.

The solution is to add one more transmon level to our toolbox. As soon as we have at least three levels, we can define a control/computational subspace and an orthogonal error space. We define our control space as the space spanned by $|g\rangle, |f\rangle$ and the error space to be $|e\rangle$. This corresponds to a new control Hamiltonian

$$H = -\frac{\alpha}{2}A^\dagger A^\dagger AA - \frac{K_C}{2}C^\dagger C^\dagger CC - \chi A^\dagger AC^\dagger C \tag{6.4}$$

$$+ \epsilon_A^*(t)|g\rangle\langle f| + \epsilon_A(t)|f\rangle\langle g| + \epsilon_C^*(t)C + \epsilon_C(t)C^\dagger. \tag{6.5}$$

that has drives of the form $|g\rangle\langle f| + h.c.$. This direct $g-f$ interaction has been previously used for fault-tolerant gates [Reinhold et al., 2020]. We discuss how to engineer this interaction, optimize controls, and measure the fidelity of a $g-f$ transmon qubit in Chapter 7.

With the three-level transmon, gate trajectories with one jump *always* end in $|e\rangle$

$$\mathcal{G}_1(t,0)\rho = \kappa_{ef}\int_0^t dt_1 U_{\text{eff}}(t,t_1)|e\rangle\langle f|U_{\text{eff}}(t_1,0)\rho U_{\text{eff}}(t_1,0)^\dagger|f\rangle\langle e|U_{\text{eff}}(t,t_1)^\dagger \tag{6.6}$$

$$= \kappa_{ef}|e\rangle\langle e| \otimes \int_0^t dt_1 \langle e|\, U_{\text{eff}}(t,t_1)|e\rangle\langle f|U_{\text{eff}}(t_1,0)\rho U_{\text{eff}}(t_1,0)^\dagger|f\rangle\langle e|U_{\text{eff}}(t,t_1)^\dagger\, |e\rangle \tag{6.7}$$

because the matrix elements $\langle n|H|e\rangle = 0$ of Eq. 6.4 are zero between $|e\rangle$ and any other qubit state, prohibiting those transitions. The resulting post-jump evolution $\langle e|U_{\text{eff}}(t,t_1)|e\rangle$ is solely generated by the cavity drives and does not drive the qubit out of $|e\rangle$. Ideally, the only path for the error state $|e\rangle$ to leak back into the control subspace is via a second jump. The probability of two jumps happening in a reasonably short gate is small, and second order. There is also the matter of no-jump backaction due to transmon relaxation, we will return to this subject in Section 6.2.2.

6.2 How to detect transmon dephasing?

Now that we have a strategy to detect transmon decay errors, we want to know if our $g-f$ strategy provides protection against dephasing. Unlike decay, the dephasing jump operator $\sqrt{2\Gamma_f^\phi}|f\rangle\langle f|$ does not couple to an orthogonal error space, but acts within our computational subspace. Dephasing during a gate

$$\mathcal{G}_1(t,0)\rho = 2\Gamma_f^\phi\int_0^t dt_1 U_{\text{eff}}(t,t_1)|f\rangle\langle f|U_{\text{eff}}(t_1,0)\rho U_{\text{eff}}(t_1,0)^\dagger|f\rangle\langle f|U_{\text{eff}}(t,t_1)^\dagger \tag{6.8}$$

leaves the transmon in a superposition of $|g\rangle, |f\rangle$. Again, measurements of $|f\rangle$ indicate certain failure, but measurements of $|g\rangle$ don't necessarily herald success. How can we ensure that measurements of $|g\rangle$ actually herald a successful gate?

6.2.1 Path-indepedent gates

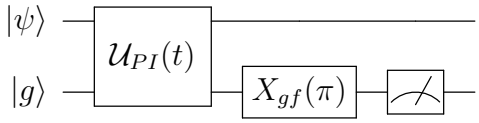
In the case of transmon relaxation, when we measure $|e\rangle$ at the end of a gate, we gain a high degree of confidence that one jump occurred. There is no such obvious flag in the case of dephasing. The transmon can yield a measurement of $|g\rangle$ at the end of a trajectory with zero or 20 jumps, there is no way for us to distinguish these trajectories. A solution to this problem is to design our gate in such a way that a final measurement of $(|f\rangle)$ $|g\rangle$ yields a (un)successful gate regardless of how many dephasing jumps occurred. This idea is known as path-independence [Ma et al., 2020].

A gate is considered path-independent if up to k jumps, a measurement of $|g\rangle$ (or some other flag state) at the conclusion of the gate heralds a successful application of that gate. Formally, we say that a gate is path-independent from $|g\rangle$ to $|g\rangle$ when

$$\langle g| \left(\sum_{p=0}^{k} \mathcal{G}_p(t,0)(|g\rangle\langle g| \otimes |\psi\rangle\langle\psi|) \right) |g\rangle \propto U_t|\psi\rangle\langle\psi|U_t^\dagger \tag{6.9}$$

where U_t is the target unitary on the cavity and the proportionality constant is a probability factor. Our goal is to satisfy this condition for one jump $k = 1$, which turns out to be quite difficult.

There is a construction that satisfies path independence to dephasing $|f\rangle\langle f|$ up to $k = \infty$. The idea is that if every incremental rotation of the transmon Bloch vector generates the same, time-independent evolution on the cavity, then the effect of dephasing will only affect the position of the transmon's Bloch vector, and won't have backaction on the cavity's evolution. Our idea is to tie the cavity's evolution to a simple transmon rotation from $|g\rangle \to |f\rangle$ followed by a reset to $|g\rangle$, which heralds a successful gate



Generally, this means that for $\mathcal{U}_{PI}$ to be path-independent, its no jump propagator

$\mathcal{G}_0(t_1,t)\rho = U_{\text{eff}}(t_1,t)\rho U_{\text{eff}}^\dagger(t_1,t)$ can be expressed as

$$U_{\text{eff}}(t_1,t) = \sum_{m,n} \xi_{mn}(t_1,t)|m\rangle\langle n| \otimes U_{mn}(t_1,t) \tag{6.10}$$

for complex valued functions $\xi_{mn}(t_1,t)$ [Ma et al., 2020].

Let's consider a brief and fairly general example of path-independence where the target cavity unitaries in the right hand side (RHS) of Eq. 6.10 are time-independent. Suppose that we can engineer the Hamiltonian

$$H = \Omega|f\rangle\langle g| \otimes U_t + \Omega|g\rangle\langle f| \otimes U_t^\dagger \tag{6.11}$$

that generates the no-jump propagator in the form of Eq. 6.10

$$\begin{aligned} U_{\text{eff}}(t_1,t) &= e^{-i(t_1-t)\left(\Omega|f\rangle\langle g|\otimes U_t + \Omega|g\rangle\langle f|\otimes U_t^\dagger\right)} && (6.12)\\ &= \cos(\Omega(t_1-t))\mathbf{I}\otimes\mathbf{I} - i\sin(\Omega(t_1-t))\left(|f\rangle\langle g|\otimes U_t + |g\rangle\langle f|\otimes U_t^\dagger\right). && (6.13)\end{aligned}$$

We don't include the transmon no-jump evolution $i|f\rangle\langle f|$ that rotates the transmon towards $|g\rangle$ here because it doesn't distort U_t. The path-independent no-jump propagator U_{eff} performs U_t on the cavity when the initial transmon state is $|g\rangle$

$$U_{\text{eff}}(\pi/(2\Omega),0)\,|g\rangle\otimes|\psi\rangle = -i\,|f\rangle\otimes U_t\,|\psi\rangle \tag{6.14}$$

and identity when the initial state is $|g\rangle$.

If a single jump occurs during the gate at time t

$$U_{\text{eff}}(\pi/(2\Omega),t)|f\rangle\langle f|U_{\text{eff}}(t,0)\,|g\rangle\otimes|\psi\rangle = -i\sin(\Omega t)(\cos(\pi/2-\Omega t)\,|f\rangle\otimes U_t\,|\psi\rangle - i\sin(\pi/2-\Omega t))\,|g\rangle \tag{6.15}$$

the final state of the transmon is entangled with a successful application of U_t or failed (identity) gate. The $g-f$ π-pulse that follows $U_{\text{eff}}(t)$ returns the transmon to the successful flag state $|g\rangle$ in the presence of no jumps. We see that the probability of a successful gate $\cos^2(\pi/2-\Omega t)$ depends on the time of the jump and decreases the later the jump occurs. When there is more than one jump, each jump effectively reverses the direction of the transmon's rotation, but the same structure and result remains

$$U_{\text{eff}}(t-t_n)|f\rangle\langle f|U_{\text{eff}}(t_n-t_{n-1})|f\rangle\langle f|\cdots|f\rangle\langle f|U_{\text{eff}}(t_1)(|g\rangle\otimes|\psi\rangle) \tag{6.16}$$

$$=\cos(\theta_n)\,|g\rangle\otimes|\psi\rangle - i\sin(\theta_n)\,|f\rangle\otimes U_t\,|\psi\rangle\,. \tag{6.17}$$

The success rate depends on the jump times where $\theta_n=\Omega\sum_{k=0}^{n}(-1)^k(t_{k+1}-t_k)$ and $t_0=0, t_{n+1}=\frac{\pi}{2\Omega}$. Remarkably, this means that dephasing jumps don't necessarily produce failed gates, but rather produce a superposition of failed and successful gates. We collapse this superposition with an error-detecting measurement at the conclusion of the gate. In this case, the failed gates perform the identity and we can reset our transmon and try again [Reinhold et al., 2020].

Engineering a Hamiltonian that produces the no-jump structure in Eq. 6.10 turns out to be very difficult. In fact, this dissertation is aware of only one gate that satisfies path independence to dephasing: selective number-dependent arbitrary phase (SNAP) gates [Krastanov et al., 2015, Ma et al., 2020, Ma et al., 2022]. The two-qubit number-conserving entangling gates in [Tsunoda et al., 2023] always fail into a detectable state after a dephasing jump, but never succeed after the jump. Both of these gates are cavity photon-number preserving and only utilize drives on the transmon ancilla and beamsplitter coupling element.

6.2.2 The equal-latitude condition

The assumption of cavity photon-number preservation simplifies our $g-f$ control Hamiltonian

$$H = -\frac{K_C}{2}C^\dagger C^\dagger C C - \chi_f |f\rangle\langle f| C^\dagger C + \epsilon^*(t)|g\rangle\langle f| + \epsilon(t)|f\rangle\langle g| \tag{6.18}$$

so that only the transmon $g-f$ drives remain. We'd like to build some intuition of what path-independence in this case means. Notice that this Hamiltonian is block-diagonal, allowing us to re-organize the composite transmon-cavity Hilbert space into a direct sum of infinitely many $g-f$ qubits conditioned on n photons in the cavity. All qubits are coupled to the same drive

$$H = \bigoplus_n \left(-\frac{K_C}{2}(n^2-n) - n\chi_f |f\rangle\langle f| + \epsilon^*(t)|g\rangle\langle f| + \epsilon(t)|f\rangle\langle g| \right) \tag{6.19}$$

with non-local dephasing operator $\prod_n |f\rangle\langle f|$ that dephases all qubits simultaneously.

So why have we transformed our coupled cavity and transmon into a countable set of qubits with a weird non-local dephasing operator? Consider the evolution of N of our qubits initialized in the state $\bigoplus_{n=0}^{N-1} |g\rangle / \sqrt{N}$. The shared drive $\epsilon(t)$ rotates each qubit leaving it in a (potential) superposition at time t

$$|\psi(t)\rangle = \frac{1}{\sqrt{N}} \bigoplus_n (a_n(t)\,|g\rangle + b_n(t)\,|f\rangle). \tag{6.20}$$

where $|a_n|^2 + |b_n|^2 = 1$ meaning that each qubit is equally represented in this superposition. Measurements $\langle Z_n \rangle$ of any single qubit yields a result between $N/(N-1)$ and $-N/(N-1)$. Now, suppose a dephasing jump occurs. The re-normalized state

after the jump is

$$\frac{|f\rangle\langle f|\,|\psi(t)\rangle}{\sqrt{\langle\psi(t)|f\rangle\langle f|\psi(t)\rangle}} = \frac{1}{\sqrt{\sum_k |b_k(t)|^2}} \bigoplus_n b_n(t)\,|f\rangle \tag{6.21}$$

and any further evolution to $t_2 > t$ produces

$$|\psi(t_2)\rangle = \frac{1}{\sqrt{\sum_k |b_k(t)|^2}} \bigoplus_n b_n(t)(a_n(t_2)\,|g\rangle + b_n(t_2)\,|f\rangle). \tag{6.22}$$

But now each qubit is not evenly represented in the resulting superposition! Instead, the qubit is now present in amount $\frac{|b_n(t)|^2}{\sum_k |b_k(t)|^2}$ which depends on the time at which the jump occurred *and* the probability of measuring $|f\rangle$. This means that not only does the dephasing jump collapse superpositions, its backaction also *reweights* the amount of each qubit in the superposition, unless all $|b_n(t)|^2$ are equal. This same effect occurs continuously for the no-jump backaction for both dephasing $|f\rangle\langle f|$ and single excitation decay $|e\rangle\langle f|$. That is, the jump and no-jump backaction is minimized only when $|b_n(t)|^2$ are equal. Each qubit can rotate about a different axis, as long as their excited state population is uniform.

Remembering that each qubit represents the transmon state conditioned on a specific cavity Fock state, this means that the jump and no-jump backaction can produce codespace logical errors or even much more harmful leakage errors. If we want the backaction of dephasing jumps to not affect our cavity state, we need the probability $|b_n(t)|^2$ of measuring $|f\rangle$ given n cavity photons to be independent of n. This is especially important for bosonic codes with codewords consisting of superpositions of Fock states. Transmon ancilla dephasing that reweights superpositions produces leakage in superposition code words, such as the kitten code word [Michael et al., 2016] $|0_L\rangle = (|0\rangle + |4\rangle)/\sqrt{2}$, unless $|b_n(t)|^2$ is independent of n. We call this condition the "equal latitude" condition.

The equal latitude condition arises naturally from the definition of the no-jump

path-independent propagator Eq. 6.10. We show this by computing the *conditional probability* $P(|f\rangle\,|n)$, which we call the *latitudes*, of measuring $|f\rangle$ when there are n photons in the cavity. We do this by computing two probabilities

$$P(|f\rangle\,|n) = \frac{P(|f\rangle \cap n)}{P(n)} = \frac{\mathrm{Tr}((|f\rangle\langle f| \otimes |n\rangle\langle n|)\rho(t))}{\mathrm{Tr}(|n\rangle\langle n|\rho(t))} \tag{6.23}$$

for the pure state $\rho(t) = U_{\text{eff}}(t_1, t)|g\rangle\langle g| \otimes |\psi\rangle\langle\psi|U_{\text{eff}}^{\dagger}(t_1, t)$. This gives a value independent of n

$$P(|f\rangle\,|n) = \frac{\mathrm{Tr}((|f\rangle\langle f| \otimes |n\rangle\langle n|)\rho(t))}{\mathrm{Tr}(|n\rangle\langle n|\rho(t))} \tag{6.24}$$

$$= \frac{|\xi_{gf}(t_1, t)|^2 |\langle n|\psi\rangle|^2}{\sum_k |\xi_{gk}(t_1, t)|^2 |\langle n|\psi\rangle|^2} \tag{6.25}$$

$$= \frac{|\xi_{gf}(t_1, t)|^2}{\sum_k |\xi_{gk}(t_1, t)|^2} \tag{6.26}$$

that is solely dependent on the transmon's trajectory. The denominator of Eq. 6.26 does not necessarily sum to one because the no-jump evolution is not trace-preserving.

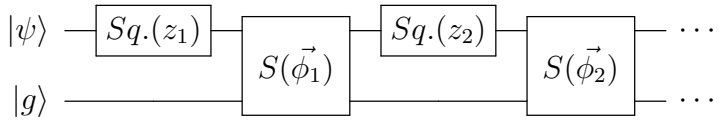
6.2.3 Path-independent SNAP

SNAP is a photon-number preserving gate on the cavity that selectively applies a complex phase to a finite number of Fock states

$$S(\vec{\phi}) = \sum_{k=0}^{k_{max}} e^{i\phi_k}|k\rangle\langle k|. \tag{6.27}$$

SNAP and displacements form a basis for cavity unitaries [Krastanov et al., 2015,Fösel et al., 2020] and is often considered as the "discrete" version of optimal control. Conveniently, SNAP directly performs $R_Z^L(\theta)$ rotations in the binomial code by applying a complex phase to $|1_L\rangle = |2\rangle$: $R_Z^L(\theta) = S(\vec{\phi})$ where $\vec{\phi}_2 = \theta$ and $\vec{\phi}_0 = \vec{\phi}_4 = 0$. Logical rotations R_X^L, R_Y^L can also be implemented with interleaved rounds of SNAP and

squeezing



Replacing displacements with squeezing maintains parity of the binomial code, preserving its error-detecting properties. The parameters $z_k, \vec{\phi}_k$ are found via numerical optimization with the same techniques described in Section 2.2.2, but where the blocks in the quantum circuit are squeezing and SNAP instead of step propagators [Fösel et al., 2020]. We have numerically found that five rounds of squeezing and SNAP applied to the $\{|0\rangle, |2\rangle, |4\rangle, |6\rangle\}$ Fock states are needed to reach discretization error of 10^{-5} for $R_X^L(\theta), R_Y^L(\theta), H^L$ gates.

Previous implementations of SNAP drove the transmon qubit with number-selective frequency combs to rotate the transmon with a different phase for each Fock state [Heeres et al., 2015, Reinhold et al., 2020]. To satisfy the equal-latitude condition, we add frequency content at both $|0\rangle$ and $|4\rangle$ with zero phase so that $P(|f\rangle|4) = P(|f\rangle|2) = P(|f\rangle|0)$. In the frame rotating with the $g-f$ transition, the SNAP Hamiltonian that implements $S(\vec{\phi})$ (neglecting Kerr for simplicity) is

$$H = -\chi_f |f\rangle\langle f| C^\dagger C + \Omega \sum_{k=0,2,4} e^{i(k\chi_f t + \phi_k)} |g\rangle\langle f| + \text{h.c.}. \tag{6.28}$$

To see the effect of each frequency component in the drive, we enter a frame rotating with the dispersive interaction as we did in Eq. 2.44. This time, the unitary transformation $\exp\left(-it\chi_f |f\rangle\langle f| C^\dagger C\right)$ produces

$$\tilde{H} = \Omega \sum_{k=0,2,4} e^{i((k-C^\dagger C)\chi_f t + \phi_k)} |g\rangle\langle f| + \text{h.c.} \tag{6.29}$$

$$= \Omega \sum_{k=0,2,4} \sum_n e^{i((k-n)\chi_f t + \phi_k)} |g\rangle\langle f| \otimes |n\rangle\langle n| + \text{h.c.}. \tag{6.30}$$

If $\chi_f \gg \Omega$, then we can neglect the off-resonant rotating terms and get a Hamiltonian

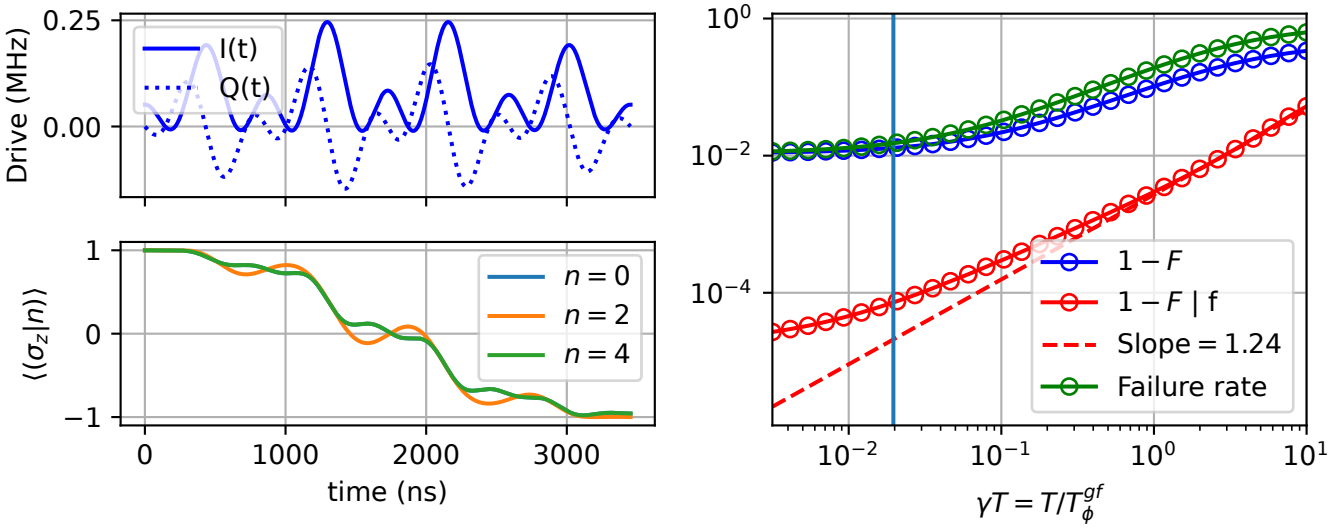


Figure 6.1: **Top left**: real $I(t)$ and imaginary $Q(t)$ parts of the control waveform $\epsilon(t)$ in Hamiltonian 6.18 (with $K_C = 0$) that implements $R(Z = \pi)$ on the binomial code with SNAP. The waveform is a sum of Gaussian envelopes with $\sigma = 1/\chi_f = 1727$ ns with total length $2\sigma = 3454$ ns modulated at frequencies $0, -2\chi_f, -4\chi_f$. **Bottom left**: the latitude of the transmon conditioned on each Fock state present in the binomial code. $\langle(\sigma_z|n)\rangle = \mathrm{Tr}(\sigma_z \otimes |n\rangle\langle n|\rho(t))/\,\mathrm{Tr}(|n\rangle\langle n|\rho(t))$ where $\rho(0) = |+y\rangle$. The $n = 0$ and $n = 4$ curves fully overlap. Clearly the waveform does not produce an equal-latitude gate, and dephasing jumps at points of non-equal latitude produce logical rotations. **Right**: simulation of the resulting gate fidelities in the presence of only dephasing decoherence $\mathcal{D}\left[\sqrt{2\Gamma_f^\phi}|f\rangle\langle f|\right]$. Here, we assume that our error detection is perfect $P(f||f\rangle) = 1, P(f||\bar{f}\rangle) = 0$ for simplicity. We plot the infidelities $1-\bar{F}$ with and without measuring $|f\rangle$ at the end of the pulse (we are not including the $g-f$ π-pulse that resets the transmon to $|g\rangle$). The error-detected fidelity is computed by measuring out the transmon in $|f\rangle$ and renormalizing. This results in replacing $\mathcal{E}[\rho]$ in Eq. 2.61 with $\rho_{\text{success}} = (|f\rangle\langle f|\mathcal{E}[\rho]|f\rangle\langle f|)/\,\mathrm{Tr}(|f\rangle\langle f|\mathcal{E}[\rho])$. For the non-error detected case, the transmon mode is traced out after simulation of $\mathcal{E}[\rho]$ so that the fidelity Eq. 6.49 is that of the resulting operation on the cavity alone. We also plot the failure rate $1 - P(|f\rangle)$. Notice that while error-detection improves fidelity by nearly two orders of magnitude at experimentally measured T_ϕ^{gf} (blue vertical line), the infidelity scales as $O\left(\left(\frac{T}{T_\phi^{gf}}\right)^{1.24}\right)$, which does not meet the quadratic error scaling that is the signature of gates that detect single jumps.

for SNAP

$$H_{\text{SNAP}} = \Omega \sum_{k=0,2,4} e^{i\phi_k} |g\rangle\langle f| \otimes |k\rangle\langle k| + \text{h.c.}. \tag{6.31}$$

$$= \Omega S(\vec{\phi})|g\rangle\langle f| + \Omega S^\dagger(\vec{\phi})|f\rangle\langle g| \tag{6.32}$$

that is in exactly the path-independent form in Eq. 6.11.

However, χ_f is typically around 1 MHz, requiring long pulses (small Ω) to satisfy the RWA condition above. If we want to approach or exceed that speed limit, we need to do additional pulse shaping and tune the values ϕ_k to account for over/undershoot resulting from the effects of the off-resonant terms [Landgraf et al., 2023]. Unfortunately, while we can correct the coherent effects of off-resonant terms, there is not an obvious way to ensure that the equal-latitude condition is met. In Fig. 6.1, we simulate the transmon trajectories of a logical $R_X^L(\pi)$ rotation on the binomial code and show that equal latitude is not satisfied, manifesting in nearly linear scaling in the error rate Γ_f^ϕ, not the quadratic error scaling we expect with true single jump detection.

6.2.4 Numerical optimization of path-independent SNAP

Since previous SNAP constructions have not satisfied the equal-latitude condition necessary for path-independence to transmon dephasing, we are forced to find a new construction. DRAG-like techniques [Gambetta et al., 2011] offer advanced pulse shaping tools, but their application to enforcing equal-latitude remains an open question to us. Instead, we will utilize the optimal control techniques reviewed in Section 2.2.2 to design pulses that are path-independent up to one dephasing jump. Optimizing in the presence of more than one jump is possible with Monte-Carlo methods, but for $p_{\text{phys}} \ll 1$ single-jump errors will be the dominant contribution to the gradient. The memory required to include multiple dephasing jumps is prohibitive for the

Monte-Carlo inspired methods we use in this section. For now, we focus on achieving path-independence up to just one jump.

Our control Hamiltonian is the same as 6.18

$$H = -\frac{K_C}{2}C^\dagger C^\dagger CC - \chi_f|f\rangle\langle f|C^\dagger C + \epsilon^*(t)|g\rangle\langle f| + \epsilon(t)|f\rangle\langle g| \tag{6.33}$$

$$= -\frac{K_C}{2}C^\dagger C^\dagger CC - \chi_f|f\rangle\langle f|C^\dagger C + I(t)\sigma_x^{gf} + iQ(t)\sigma_y^{gf} \tag{6.34}$$

where $I(t) = \Re(\epsilon(t))$ and $Q(t) = \Im(\epsilon(t))$. To make it easier to experimentally measure the fidelity of our SNAP gates, we choose to optimize logical Z rotations on the binomial code as defined at the beginning of 6.2.3. In particular, we optimize pulses to realize gates $|g\rangle\langle g| \otimes G_k$ for logical gates G_k in the S^L group $G_S = \{\mathbf{I}, R_Z^L(\pi), R_Z^L(\pi/2), R_Z^L(-\pi/2)\}$ (note that $R_Z^L(\pi/2)$ is often referred to as S^L, in which case $G_S = \{\mathbf{I}, R_Z^L(\pi), S^L, (S^L)^\dagger\}$) to perform the logical unitary when the transmon begins and ends in $|g\rangle$.

The first step when numerically optimizing pulse controls is to choose a cost function. The evolution we want to simulate, dephasing, produces non-unitary evolution, so we cannot use the unitary cost functions in Section 2.2.2. Instead, we must use the general fidelity function Eq. 2.58 that quantifies the fidelity of a channel. Since our logical space has dimension[1] $d_L = 2$, we use the channel fidelity for a qubit Eq. 2.61. This requires that we compute the evolution $\mathcal{U}(|g\rangle\langle g| \otimes \rho_L)$ for the $d_L^2 = 4$ logical code words $\rho_L \in \{|+\rangle\langle +|, |-\rangle\langle -|, |0\rangle\langle 0|, |1\rangle\langle 1|\}$. We assume that the transmon is always initialized in $|g\rangle$ at the beginning of a gate and again call our target unitary on the cavity mode U_t. Four states are needed in this case to span the image of any $\mathcal{U}$, but it can be possible to use fewer initial states depending on the structure of the channel's errors. This can be especially useful in multi-qubit simulations.

1. The dimension d of the underlying Hilbert space on which the code words are defined can be higher.

To optimize gates up to one dephasing jump, we want to simulate evolution produced by the channel $\mathcal{U}$ up to one dephasing jump. In this case, we write the fewer-than-two-jumps channel $\mathcal{U}^{(1)}$ as the sum of the no-jump and one-jump propagators by truncating the Dyson series Eq. 5.2

$$\mathcal{U}^{(1)}\rho = \frac{1}{\mathcal{N}}\left(\mathcal{G}_0(t,0)\rho + \mathcal{G}_1(t,0)\rho\right) \tag{6.35}$$

$$= \frac{1}{\mathcal{N}}\left(U_{\text{eff}}\rho U_{\text{eff}}^\dagger + 2\Gamma_f^\phi \int_0^t dt_1 U_{\text{eff}}(t,t_1)|f\rangle\langle f|U_{\text{eff}}(t_1,0)\rho U_{\text{eff}}(t_1,0)^\dagger|f\rangle\langle f|U_{\text{eff}}(t,t_1)^\dagger\right) \tag{6.36}$$

for U_{eff} defined in Eq. 5.5 and $\mathcal{G}_1(t,0)$ defined in Eq. 6.8. We need a normalization factor $\mathcal{N} = \text{Tr}\left(\mathcal{G}_0(t,0)\rho + \mathcal{G}_1(t,0)\rho\right)$ such that $\text{Tr}(\mathcal{U}^{(1)}\rho) = 1$ because the truncated Dyson series may not preserve trace. $\mathcal{N}$, which depends on ρ, is the probability that fewer than two jumps occur during the evolution of ρ.

The first term in Eq. 6.36 is easy to evaluate by computing the no-jump evolution U_{eff} from the step propagators in Eq. 2.48. The one-jump propagator in the second term looks more intimidating, but is vastly simplified when ρ is a pure state $\rho_\psi = |\psi\rangle\langle\psi|$. Then, the final density matrix is an average

$$\mathcal{U}^{(1)}\rho_\psi = \frac{1}{\mathcal{N}}\left(U_{\text{eff}}|\psi\rangle\langle\psi|U_{\text{eff}}^\dagger + 2\Gamma_f^\phi \int_0^t dt_1 |\psi(t_1)\rangle\langle\psi(t_1)|\right) \tag{6.37}$$

over the jump time t_1 of all trajectories

$$|\psi(t_1)\rangle = U_{\text{eff}}(t,t_1)|f\rangle\langle f|U_{\text{eff}}(t_1,0)\,|\psi\rangle\,. \tag{6.38}$$

We do not renormalize the states $|\psi(t_1)\rangle$, as the norm represents the probability of a trajectory with a jump at time t_1 occurring.

Computing η trajectories for d_L^2 state vectors of physical dimension d requires $O(\eta d_L^2 d^3)$ operations. Propagating d_L^2 density operators each of dimension $d \times d$ is

much more costly, requiring $O(d_L^2 d^6)$ operations [2] [Abdelhafez et al., 2019]. To accurately approximate density matrix evolution with trajectories, η can be as small as $\eta = 10$ for systems with low error rates [Abdelhafez et al., 2019]. This provides a significant advantage over the density operator method, especially for bosonic systems with $d \gg 2$. Minimizing required resources not only benefits runtime; it also lowers the memory overhead of using automatic differentiation techniques to compute gradients [van Merriënboer et al., 2019].

In fact, computing the expectation value of any operator after evolution of a pure state $\rho_\psi = |\psi\rangle\langle\psi|$ in a non-unitary channel can be translated to an average over trajectories

$$\mathrm{Tr}(A\mathcal{U}\rho_\psi) = \overline{\langle\varphi|A|\varphi\rangle}. \tag{6.39}$$

The overbar indicates an average over all of the final states $|\varphi\rangle$ generated by standard Monte-Carlo trajectory techniques. The uncertainty in the average is determined by the number of trajectories over which one averages. Chapter 4 of [Wiseman and Milburn, 2014] is an excellent reference for computing Monte-Carlo trajectories. We have implemented this trajectory method using techniques that reduce the sampling cost [Abdelhafez et al., 2019], but do not utilize it here as we are focused on first tackling the one-jump trajectories.

Instead of randomly sampling the jump times as in Monte-Carlo methods, we evenly sample the t_1. This converts the integral in Eq. 6.36 that averages over all jump times t_1 into a sum

$$\mathcal{U}^{(1)}\rho_\psi = \frac{1}{\mathcal{N}}\left(U_{\text{eff}}|\psi\rangle\langle\psi|U_{\text{eff}}^\dagger + 2\Gamma_f^\phi dt \sum_{k=0}^{N-1} |\psi(kdt)\rangle\langle\psi(kdt)|\right) \tag{6.40}$$

2. This is similar to the method for finding $\mathcal{U}^{(1)}$ itself, which requires propagating d^2 physical basis states.

where dt is the timestep used in the step propagators. The total pulse length is $t = Ndt$. Although we evenly sample jumps in time, but the no-jump evolution of U_{eff} correctly encodes the probability of a jump (or no jump) at time t_1 in the norm of that trajectory. Computing the average value of A with this state

$$\text{Tr}(A\mathcal{U}^{(1)}\rho_\psi) = \frac{1}{\mathcal{N}}\left(\text{Tr}\left(AU_{\text{eff}}|\psi\rangle\langle\psi|U_{\text{eff}}^\dagger\right) + 2\Gamma_f^\phi dt\sum_{k=0}^{N-1}\text{Tr}\left(A|\psi(kdt)\rangle\langle\psi(kdt)|\right)\right) \tag{6.41}$$

$$= \frac{1}{\mathcal{N}}\left((\langle\psi|U_{\text{eff}}^\dagger)A(U_{\text{eff}}|\psi\rangle) + 2\Gamma_f^\phi dt\sum_{k=0}^{N-1}\langle\psi(kdt)|A|\psi(kdt)\rangle\right) \tag{6.42}$$

$$= \frac{1}{\mathcal{N}(\Psi(|\psi\rangle))}\sum_{|k\rangle\in\Psi(|\psi\rangle}\langle k|A|k\rangle \tag{6.43}$$

reveals that it is equal to averaging (summing) over the set trajectories

$$\Psi(|\psi\rangle) = \{U_{\text{eff}}|\psi\rangle\}\cup\left\{\sqrt{2\Gamma_f^\phi dt}\,|\psi(kdt)\rangle\right\}_k. \tag{6.44}$$

which a function of the initial pure state $|\psi\rangle$. It's easy now to explicitly write down the normalization factor

$$\mathcal{N}(\Psi(|\psi\rangle)) = \sum_{|k\rangle\in\Psi}|\langle k|k\rangle|^2 \tag{6.45}$$

to explicitly indicate its dependence on the set of trajectories Ψ.

Generating these trajectories for each of the four initial states $\Phi = \{|g\rangle\otimes|+\rangle, |g\rangle\otimes|-\rangle, |g\rangle\otimes|0\rangle, |g\rangle\otimes|1\rangle\}$ allows us to compute each trace in Eq. 2.61 sampling from

only one jump trajectories. Putting it all together, the average gate fidelity is

$$\bar{F}_g(U_{\text{target}}, \mathcal{U}^{(1)}\rho) = \frac{1}{4}\sum_{|\phi\rangle\in\Phi} \text{Tr}\left(U_{\text{target}}|\phi\rangle\langle\phi|U_{\text{target}}^\dagger \mathcal{U}^{(1)}(|\phi\rangle\langle\phi|)\right) \tag{6.46}$$

$$= \sum_{|\phi\rangle\in\Phi}\sum_{|k\rangle\in\Psi(|\phi\rangle)} \frac{\langle k|\, U_{\text{target}}|\phi\rangle\langle\phi|U_{\text{target}}^\dagger\, |k\rangle}{4\mathcal{N}(\Psi(|\phi\rangle))} \tag{6.47}$$

$$= \sum_{|\phi\rangle\in\Phi}\sum_{|k\rangle\in\Psi(|\phi\rangle)} \frac{|\,\langle k|U_{\text{target}}|\phi\rangle\,|^2}{4\mathcal{N}(\Psi(|\phi\rangle))} \tag{6.48}$$

which is an average over trajectories and initial states similar in form to the incoherent state transfer fidelity Eq. 2.57. The difference here is that our set of initial states spans all density matrices, so the information lost when taking the magnitude of a complex number is restored by using a larger basis.

Notice that the fidelity computed here is the channel fidelity to the target unitary U_{target} defined on both the transmon and cavity space. To find the average gate fidelity on the cavity, we can perform a partial trace of the qubit and compute the average gate fidelity to a cavity target unitary U_t

$$\bar{F}_g(U_t, \mathcal{U}) = \frac{1}{4}\sum_{\rho\in\{|+\rangle\langle+|,|-\rangle\langle-|,|0\rangle\langle 0|,|1\rangle\langle 1|\}} \text{Tr}\left(U_t\rho^\dagger U_t^\dagger\, \text{Tr}_A\left(\mathcal{U}(|g\rangle\langle g|\otimes\rho)\right)\right) \tag{6.49}$$

where Tr_A indicates a partial trace over the mode with A annihilation operators, the transmon. This represents the case where we reset the transmon at the end of each gate. We utilize this formula to compute the non-error-detected fidelity in Fig. 6.1.

However, taking a partial trace doesn't help us if we want to optimize an error-detected gate. In that case, we need to optimize a slightly different quality that is the fidelity between a cavity operation and target U_t conditioned on the measurement of a transmon state.

Conditional fidelity

The previously described technique gives us a recipe for computing the unconditional fidelities of gates with zero or one jumps. It is difficult to optimize a gate that is path-independent to dephasing, but always successfully ends with the transmon in $|g\rangle$. What we want for error detection is to optimize the path-independence of a gate via the resulting fidelity when the error detection flags a gate as successful, as in Section 5.3.

In Chapter 5, we calculated the fidelity of error-detected gates after successful projection into a particular transmon state that heralds a successful gate. In particular, the state at the output of an error-detected channel $\mathcal{U}$ is

$$\rho_{\text{success}} = \frac{M_g \mathcal{U} \rho M_g}{\text{Tr}\left(M_g \mathcal{U} \rho M_g\right)} \tag{6.50}$$

from Eq. 5.16 for a success measurement operator M_g. To define the conditional average gate fidelity, we replace the output of the channel $\mathcal{U}^{(1)}(|\psi\rangle\langle\psi|)$ in Eq. 6.46 with the renormalized output of the error-detecting measurement

$$\bar{F}_g(U_{\text{target}}, \mathcal{U}^{(1)}) = \frac{1}{4} \sum_{|\phi\rangle \in \Phi} \frac{\text{Tr}\left(U_{\text{target}}|\phi\rangle\langle\phi|U_{\text{target}}^\dagger M_g \mathcal{U}^{(1)}(|\phi\rangle\langle\phi|) M_g\right)}{\text{Tr}\left(M_g \mathcal{U}^{(1)}|\phi\rangle\langle\phi| M_g\right)} \tag{6.51}$$

$$= \sum_{|\phi\rangle \in \Phi} \sum_{|\psi\rangle \in \Psi(|\phi\rangle)} \frac{1}{4\mathcal{N}(\Psi)} \frac{|\langle\psi| M_g U_t |\phi\rangle|^2}{\langle\psi| M_g M_g |\psi\rangle}. \tag{6.52}$$

where U_t is the target unitary on the cavity mode.

We have to be careful when optimizing the conditional fidelity above. Left to its own devices the optimizer will maximize the ratio in Eq. 6.52, but may do so with small numerator (unconditional fidelity) and denominator (success rate). To push the optimizer towards a solution with near-unity numerator and denominator, we slowly phase in the denominator as a function of the numerator. This ramp function does

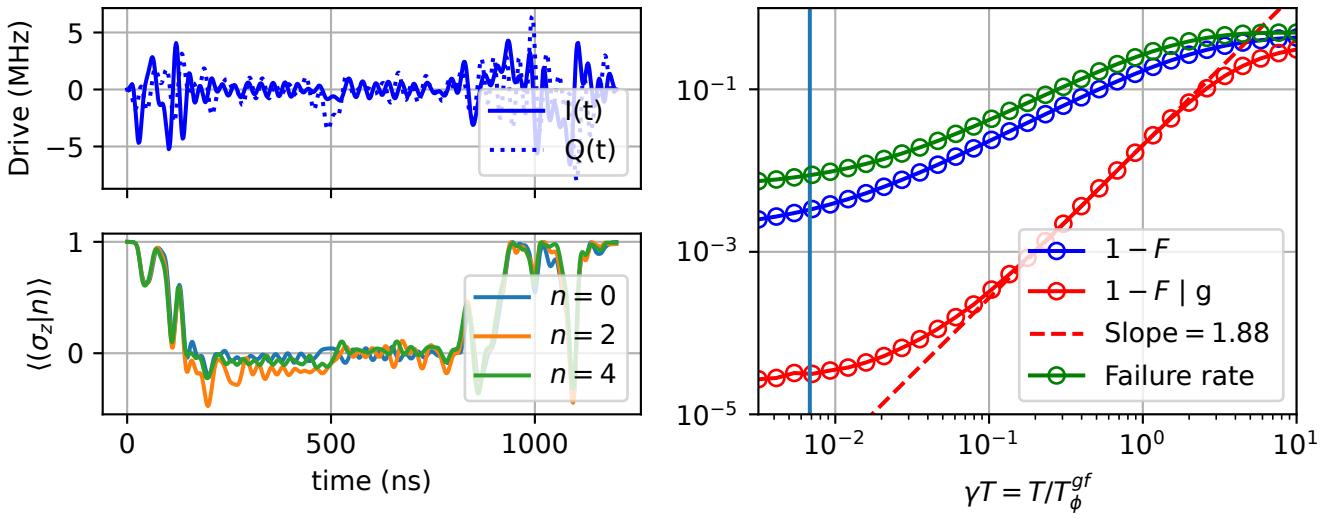


Figure 6.2: This figure has exactly the same format as Fig. 6.1. **Top left**: real $I(t)$ and imaginary $Q(t)$ parts of the control waveform $\epsilon(t)$ in Hamiltonian 6.18 that implements $R(Z = \pi)$ on the binomial code with SNAP. The Hamiltonian parameters for the optimization are listed in Table 7.1. **Bottom left**: the latitude of the transmon conditioned on each Fock state present in the binomial code. The latitudes closely overlap for the entire gate duration. **Right**: simulation of the resulting gate fidelities using Eq. 2.58 in the presence of only dephasing decoherence $\mathcal{D}\left[\sqrt{2\Gamma_f^\phi}|f\rangle\langle f|\right]$. With numerical optimization, the infidelity now scales as $O\left(\left(\frac{T}{T_\phi^{gf}}\right)^{1.88}\right)$, is substantially closer to quadratic scaling than the manually constructed waveform in Fig. 6.1.

not need to be complicated. In fact, we use a simple rectified linear unit (ReLU in machine learning parlance) as a ramp. ReLU ramps the denominator from 1 to $\langle\psi|M_g M_g|\psi\rangle$ between two chosen average values $p_{\text{start}}, p_{\text{end}}$ of the numerator

$$\sum_{|\phi\rangle\in\Phi}\sum_{|\psi\rangle\in\Psi(|\phi\rangle)}\frac{1}{4\mathcal{N}(\Psi)}\frac{|\langle\psi|\,M_g U_t\,|\phi\rangle|^2}{R\left(\max_{|k\rangle\in\Psi(|\phi\rangle}(|\langle\psi|\,M_g U_t\,|k\rangle|^2),\langle\psi|M_g M_g|\psi\rangle\,,p_{\text{start}},p_{\text{end}}\right)}. \tag{6.53}$$

The ramp function R is defined as

$$R(x,y,p_{\text{start}},p_{\text{end}}) = 1-(1-y)\min\left(\max\left(0,\frac{x-p_{\text{start}}}{p_{\text{end}}-p_{\text{start}}}\right),1\right) \tag{6.54}$$

We have found this to be an essential ingredient in finding high-fidelity path-independent SNAP gates with high success rate and generally choose $p_{\text{start}} \sim 0.85, p_{\text{end}} \sim 0.90$. Fig. 6.2 shows an example numerically optimized SNAP gate whose fidelity scales as $O\left(\left(T/T_\phi^{gf}\right)^{1.88}\right)$, which is nearly quadratic.

Gates implemented with optimal control methods are infamous for having waveforms and state trajectories that make little intuitive sense. At first glance, the waveform in Fig. 6.2 seems to have little structure, but a closer analysis in Fig. 6.3 of a $R_Z^L(-\pi/2)$ logical rotation shows that our optimizer produces a pulse with decipherable structure. The optimizer the dispersive interaction and fast, unselective π-pulses to impart a differential phase between the logical code words $|0_L\rangle, |1_L\rangle$. Remarkably, the gate has the same structure as a manual construction utilizing the "exponential gadget" technique [Tsunoda et al., 2023] that provides a design for logical $R_Z^L(\theta)$ and $R_{ZZ}^L(\theta)$ gates across many codes. These error-detectable gates are also designed to detect up to one dephasing jump using a $g-f$ transmon, so it is exciting and encouraging that our numerical technique finds the same construction. In the presence of multiple jumps, the exponential gadget construction does not always flag a failed gate (see Appendix I of [Tsunoda et al., 2023]). Although not explored here, optimization

in the presence of multiple jumps could reveal new discrete constructions that detect multiple errors. We can also use our numerical techniques to explore new constructions other than the "exponential gadget" in the Schwinger picture of [Tsunoda et al., 2023].

6.3 Conclusion

The techniques described in this chapter for designing error-detected gates focus heavily on errors in the transmon mode. This is intentional since its coherences are so much shorter than the cavity mode's. The two main errors occuring in the transmon have vastly different solutions. The first, excitation decay, can be detected up to one error by expanding the code distance of the transmon control space. This is accomplished by adding a third level $|f\rangle$ so that decay produces a transition to a detectable orthogonal error space $|e\rangle$.

As we've seen, transmon dephasing is much more of a challenge than decay. Path-independence describes a framework in which the resulting gate at the end of some evolution depends solely upon the concluding measurement result of the transmon. This perfectly fits our error-detection goals, but does not prescribe a method to find a path-independent construction for any gate. One of the only known path-independent gates is SNAP, provided that a rather strict Hamiltonian condition $\Omega \ll \chi$ is met. Our current systems generally don't meet this requirement for the speeds at which we want to operate the gate, which break the equal-latitude condition required for path-independence.

We tackle this problem by formulating a version of optimal control that optimizes a pulse over trajectories that contain up to one jump. Our software pushes us closer to quadratic scaling in the dephasing error rate, suggesting that we can detect nearly all deleterious single jump trajectories. We have tried to use this approach to find other

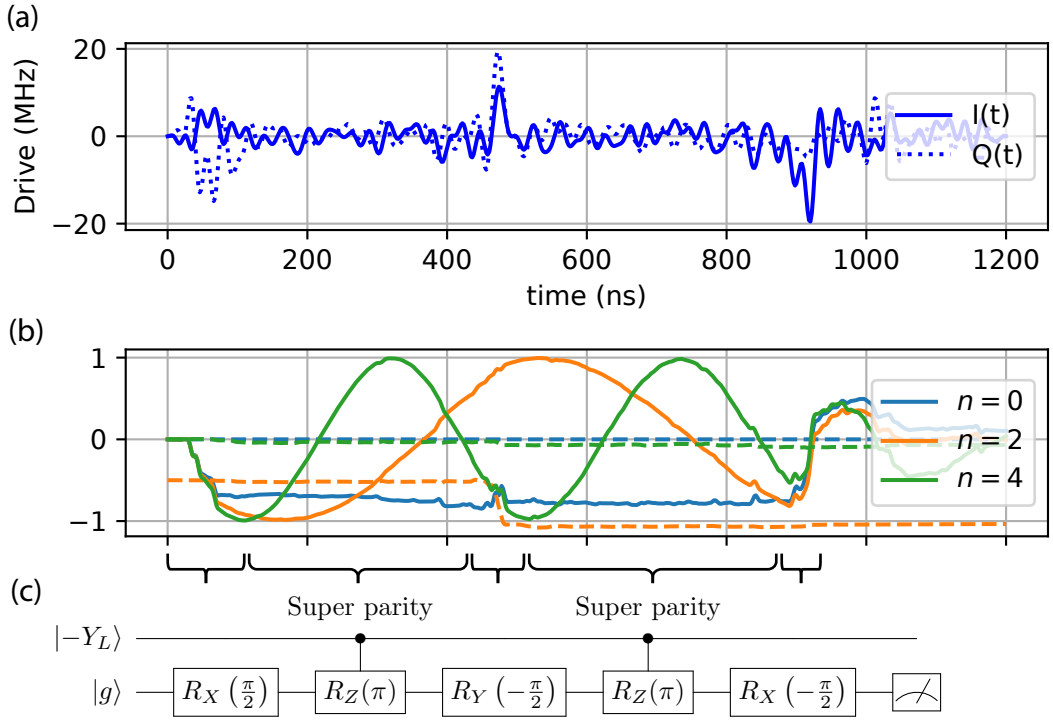


Figure 6.3: Waveform (a) and state trajectories for a $R_Z^L(-\pi/2)$ logical rotation on the binomial code. In (b), we plot the conditional transmon σ_y expectation values $\langle(\sigma_y|n)\rangle = \mathrm{Tr}(\sigma_y \otimes |n\rangle\langle n|\rho(t))/\,\mathrm{Tr}(|n\rangle\langle n|\rho(t))$ (solid lines) for the three Fock states $n = 0, 2, 4$ present in the code words. These expectation values indicate evolution along the equator of each Bloch sphere conditioned on n photons, but do not contain global phases. To see the evolution of the phase between each photon-number Bloch sphere, we look at the phase of the $|g\rangle\langle g| \otimes |n\rangle\langle 0|$ density matrix elements $\arg\left(\mathrm{Tr}\left(|g\rangle\langle g| \otimes |n\rangle\langle 0|\rho(t)\right)\right)/\pi$ (dashed lines). With both of these quantities, we identify five different, discrete stages of the gate with the circuit (c). First, there is a short, high-amplitude σ_x rotation bringing the three Bloch vectors onto the equator. This is followed by idle evolution under the dispersive interaction of duration $\sim \chi_f/(4\pi) = 432$ ns implementing a super-parity selective $R_Z(\pi)$ rotation on the three Bloch spheres. This rotation anti-aligns the Bloch vectors of $|0\rangle, |4\rangle$ to $|-Y\rangle$ and $|2\rangle$ to $|+Y\rangle$. Odd-numbered Fock states in the error space are aligned with $|\pm X\rangle$. The main trick occurs with the following $R_Y(-\pi/2)$ rotation. This rotation doesn't rotate any of the individual code space Bloch vectors, but applies a global, differential phase $\theta = -\pi/2$ between states entangled with $|-Y\rangle$ and $|+Y\rangle$. That is, $R_Y(\theta)(|-Y\rangle \otimes (|0\rangle + |4\rangle) + |+Y\rangle \otimes |2\rangle) = e^{i\theta}\,|-Y\rangle \otimes (|0\rangle + |4\rangle) + |+Y\rangle \otimes |2\rangle$. This rotation is at the heart of the exponential gadget and is where the tunable phase θ is applied [Tsunoda et al., 2023]. The next stage of the gate is another number-selective $R_Z(\pi)$ to re-align all Bloch vectors at $|-Y\rangle$. The final unselective $R_X(-\pi/2)$ returns all code space Bloch vectors to $|+Z\rangle = |g\rangle$. Bloch vectors corresponding to the odd-numbered Fock states end on the equator. The measurement at the end then detects up to one transmon dephasing or decay jump during the gate, as well as the presence of odd Fock states with low efficiency.

gates that are path-independent up to one dephasing jump, but so far have failed to find path-independent Fock state creation or R_X^L, R_Y^L rotations on the binomial code. It is likely that an additional resource is needed to unlock path-independent implementations, such as dynamic control of χ.

Chapter 7

Controlling the transmon qutrit

Typically, only the lowest two transmon energy eigenstates are used as a qubit. At Yale, we use these two levels to assist with cavity state preparation, manipulation, and measurement. Other approaches use the transmon qubit directly for quantum simulation [Kandala et al., 2019, Kim et al., 2023] and error correction [Andersen et al., 2020, Acharya et al., 2023, Jurcevic et al., 2021]. This $g-e$ transmon qubit is the easy to calibrate and control, but alone does not have any inherent error detecting properties.

In fact, there has been a significant amount of effort in ensuring that the transmon operates as a two-level $g-e$ system. Fast $g-e$ qubit gates produce leakage to $|f\rangle$ and higher levels which can be mitigated using the DRAG-like techniques discussed in Section 2.2.1. These techniques reduce the leakage, but do not completely eliminate leakage from gates and idle heating. This type of leakage is an unrecoverable error in surface codes, encouraging the development of leakage reduction units (LRUs) that quickly reset the transmon back into the $g-e$ qubit subspace [Vittal et al., 2023, McEwen et al., 2021, Lacroix et al., 2023, Battistel et al., 2021]. Many of these methods are autonomous to avoid the delay incurred by the latency of measurement and feedback.

However, the higher energy levels of the transmon are not only a nuisance spoiling $g-e$ qubits, but in some applications have real advantages. End-of-line transmon measurements can be improved by first exciting $|e\rangle$ to a higher level so that any energy decay during the readout will not confuse $|g\rangle$ with $|\bar{g}\rangle$ [Elder et al., 2020]. We can define $g-f$ erasure qubits where any measurement of $|e\rangle$ detects a decay error [Kubica et al., 2022].

Any gate construction where our goal is to use three transmon levels to detect single decay jumps, such as in Chapter 6, requires a direct coupling between the $|g\rangle$ and $|f\rangle$ states of the transmon. This type of interaction has been used previously in fault-tolerant logical error detection and logical gates [Rosenblum et al., 2018, Reinhold et al., 2020]. In this chapter, we derive another way to engineer this coupling that is well-suited for use in optimal control. We optimize $g-f$ qubit gates and logical gates on a binomial qubit and measure their fidelities with randomized benchmarking. The techniques developed here are applied in the following chapter in our demonstration of an error-detected gate.

7.1 Engineering the $g-f$ interaction

Conventionally, control over the three level transmon relies on individual control of the $g-e$ and $e-f$ transitions using the Hamiltonian in Eq. 2.36 repeated here:

$$H = \Delta|e\rangle\langle e| + (2\Delta - \alpha)|f\rangle\langle f| + \epsilon^*(t)\left(|g\rangle\langle e| + \sqrt{2}|e\rangle\langle f|\right) + \text{h.c.}. \tag{7.1}$$

Δ is the drive detuning from ω_A, so drives with $\Delta = 0$ activate $g-e$ transitions, while drives with $\Delta = -\alpha$ activate $e-f$ transitions. These drives must have bandwidth less than α to avoid spuriously driving the other transition. Resonantly driving these transitions enables full control the the $g-e-f$ transmon, but there is not an obvious sequence that prepares a superposition $|g\rangle + |f\rangle$ without intermediate population in

$|e\rangle$. If decay occurs at this point, it cannot reliably be detected as it falls back into the $g-f$ computational space. We need a direct $|g\rangle\langle f| + h.c.$ interaction.

7.1.1 Raman drives

The three level, two drive case mentioned above was first examined in optical systems. Stimulated Raman adiabatic passage (STIRAP) minimizes intermediate leakage (to $|e\rangle$, in our language) during population transfer in a three-level system by driving each transition detuned from resonance in a counter-intuitive order [Gaubatz et al., 1988,Unanyan et al., 1998,Fleischhauer et al., 1999]. In our notation, this corresponds to activating the $g-e$ coupling after the $e-f$ coupling to realize $|g\rangle \rightarrow |f\rangle$. Laser pulse-shaping is generally difficult, but simple Gaussian waveforms are sufficient for STIRAP. Extensions to superconducting qubits [Zheng et al., 2022] take advantage of microwave arbitrary waveform generators to extract additional performance.

In this section, we discuss the transmon Raman transition in the context of constructing arbitrary gates, not just $|g\rangle \leftrightarrow |f\rangle$ population inversion. We enact the two photon $g-f$ interaction using drive frequencies that sum to the energy difference $2\omega_A - \alpha$ between $|g\rangle$ and $|f\rangle$. That is, $\epsilon(t)$ has two carrier frequencies $\omega_A - \Delta, \omega_A - \alpha + \Delta$ and two envelopes $\epsilon_1(t), \epsilon_2(t)$

$$\epsilon(t) = \epsilon_1(t)e^{i\Delta t} + \epsilon_2(t)e^{-i(-\alpha+\Delta)t}. \tag{7.2}$$

There are no restrictions on the sign of Δ, but its magnitude will determine the interaction rate and leakage rate to $|e\rangle$. Here, we choose the sign of $\Delta = \omega_A - \omega_1$ so that Δ represents the amount the $g-e$ drive is red detuned from resonance, which corresponds to the amount the $e-f$ drive is blue detuned. We illustrate the spectral location of these drives in Fig. 7.1.

We can find the effect of these drives by plugging in $I(t) = \Re(\epsilon(t)), Q(t) = \Im(\epsilon(t))$

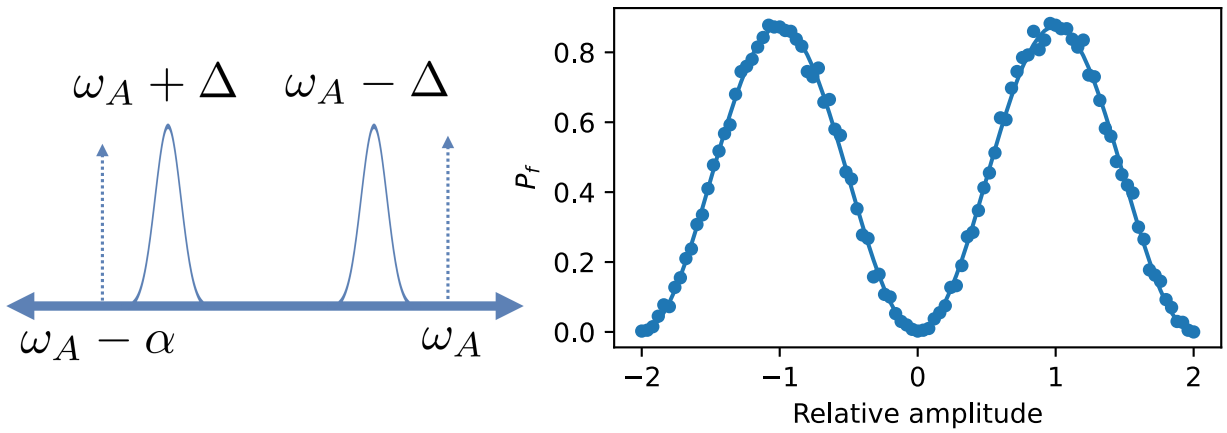


Figure 7.1: **Left**: to activate the $g-f$ two photon transition, we place two drives detuned from the $g-e$ and $e-f$ transitions by an equal but opposite amount. The spectral content of each pulse needs to be less than Δ to avoid overlap with the single photon transitions. **Right**: a power Rabi experiment where we vary only the amplitude of the Gaussian drive at $\omega_A - \alpha + \Delta$. The data (dots) are fitted to $A\sin(Bx+C)+D$ (solid line). B corresponds to the amplitude required to perform a $R_X(\pi)$ rotation of the $g-f$ qubit.

into the DRAG Hamiltonian Eq. 2.40

$$H_{\text{DRAG}} = (I_1(t) + \cos((\alpha-\Delta)t)I_2(t))\sigma_{ge}^x - \frac{2I_1(t)^2 + I_2(t)^2}{\sqrt{2}\alpha} - \frac{4I_1(t)I_2(t)\cos((\alpha-\Delta)t)}{\sqrt{2}\alpha}\sigma_{gf}^x \tag{7.3}$$
$$+ \left(2\Delta - \alpha - \frac{4I(t)^2}{\alpha}\right)|f\rangle\langle f| + \left(\Delta - \frac{2I(t)^2}{\alpha}\right)|e\rangle\langle e| + \left(Q - \frac{\dot{I}(t)}{\alpha}\right)(\sigma_{ge}^y + \sqrt{2}\sigma_{ef}^y).$$

neglecting terms that rotate at $2(\alpha-\Delta)$. We've only expanded one of the $I(t)^2$ terms to limit the mess. As before, H_{DRAG} contains undesired $g-e$ and $e-f$ rotations, off-resonant $g-f$ rotations, and now the desired on-resonant σ_{gf}^x term. The term that survives further RWAs is $4I_1(t)I_2(t)\cos((\alpha-\Delta)t$ which is detuned by only Δ. Notably, the Rabi rate is the product of both drive envelopes $I_1(t)I_2(t)$. This is rather convenient as we separate the two drives into a "pump" mode with a simple Gaussian envelope and "control" mode that contains different frequency components for selective rotations [Rosenblum et al., 2018] and phases for SNAP [Reinhold et al., 2020]. We plot a power Rabi in Fig. 7.1 that demonstrates linear amplitude squaring

in each drive.

There do exist a couple of drawbacks with this Raman scheme. The first is that the rate of unselective $g-f$ pulses is set by Δ which is typically substantially smaller than α. As one decreases the duration of a $g-f$ π-pulse its bandwidth increases. If the bandwidth of this pulses exceeds Δ, the spectrum of the waveform can overlap with the $g-e$ or $e-f$ transitions, producing leakage to $|e\rangle$ and spoiling a gate's decay detecting potential. It's likely that DRAG techniques to minimize leakage are applicable, but we do not explore those here. Short of non-trivial pulse shaping, one should ask: is there an optimal choice of Δ where leakage is minimized? Yes! We will return to this case shortly, but it turns out that $\Delta = \alpha/2$ places both drives spectrally equidistant between the $g-e$ and $e-f$ transitions.

The second drawback of the Raman scheme is that there does not exist a frame in which the only explicit time dependence in the Hamiltonian is from baseband envelopes $\epsilon_1(t), \epsilon_2(t)$. This is a problem for optimal control, where we want to avoid simulating the evolution of Hamiltonians with fast oscillating terms $e^{i\omega t}$. We always perform optimal control in the rotating frame that minimizes these terms to avoid accumulating error when we approximate the evolution with step-propagators. There is also the practical aspect of ensuring that time step of the simulation dt is small enough to accurately sample the fast oscillations.

7.1.2 The $\Delta = \alpha/2$ case

The case when the detuning $\Delta = \alpha/2$ is worth treating on its own. In this case, both of the Raman drives have the same frequency, so we will continue with only one

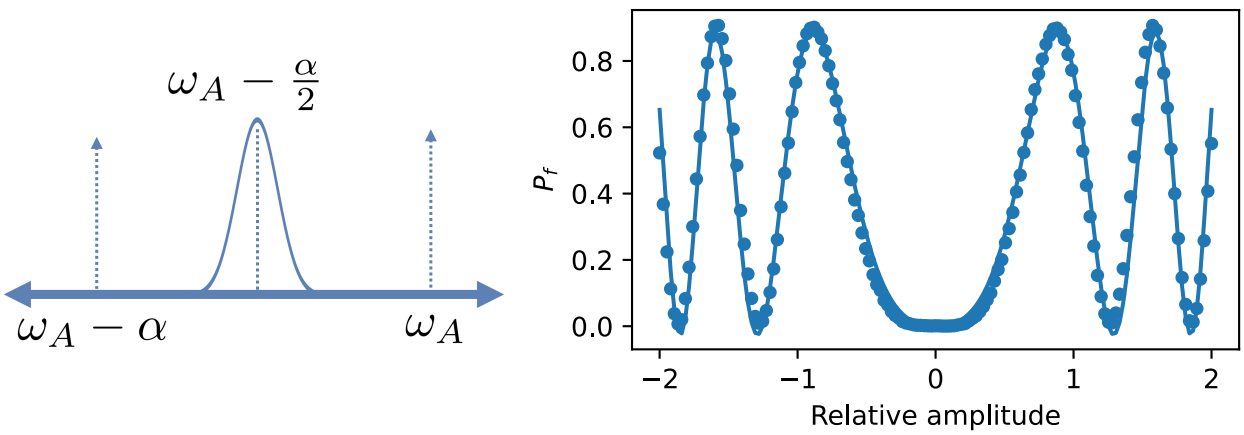


Figure 7.2: **Left**: to activate the $gf/2$ two photon transition, we place a drive at $\omega_A - \alpha/2$ between the single photon $g-e$ and $e-f$ transitions. This provides spectral separation between the $gf/2$ transition and either single photon transition of $\alpha/2$. **Right**: a power Rabi experiment where we vary the amplitude of a Gaussian drive and observe that the oscillations depend on the square of the drive amplitude. The data (dots) are fitted to $A\sin(Bx^2+C)+D$ (solid line). B corresponds to the amplitude required to perform a $R_X(\pi)$ rotation of the $g-f$ qubit.

envelope $\epsilon(t)$. Re-examining the DRAG Hamiltonian Eq. 2.40

$$\begin{aligned} H_{\text{DRAG}} =& I(t)\sigma_{ge}^x - \frac{I(t)^2}{\sqrt{2}\alpha}\sigma_{gf}^x + \left(2\Delta - \alpha - \frac{4I(t)^2}{\alpha}\right)|f\rangle\langle f| \\ &+ \left(\Delta - \frac{2I(t)^2}{\alpha}\right)|e\rangle\langle e| + \left(Q - \frac{\dot{I}(t)}{\alpha}\right)(\sigma_{ge}^y + \sqrt{2}\sigma_{ef}^y). \end{aligned} \tag{7.4}$$

We want to choose Δ and $Q(t)$ to bring the $g-f$ coupling into resonance and cancel the residual $g-e$ and $e-f$ couplings and Stark shifts while leaving other $g-e$ terms far off-resonant. We eliminate the deleterious terms by choosing

$$Q(t) = \frac{\dot{I}(t)}{\alpha} \tag{7.5}$$

$$\Delta(t) = \frac{\alpha}{2} + \frac{2I(t)^2}{\alpha} \tag{7.6}$$

which are nearly the same choices we make to minimize leakage to $|f\rangle$ for $g-e$ drives. We call this technique for engineering the $g-f$ interaction the "$gf/2$" method.

A substantial difference between the $gf/2$ approach and the previously discussed

two drive scheme is that the $gf/2$ Rabi rate is the drive amplitude squared $I(t)^2$. We demonstrate amplitude squaring with a simple power Rabi experiment in Fig. 7.2. With two drives, we are able to assign one as a pump and use the second to selectively drive transitions. Now, the amplitude squaring works as a form of two-wave mixing generating additional frequency components. For example, a selective $g-e$ π-pulse with frequency content at $0, -2\chi, -4\chi$ implements a parity check. However, the $gf/2$ amplitude squaring term $I(t)^2$ mixes the frequency components present in a selective π-pulse and changes the states to which the pulse is selective. For example, a drive $I(t) = 1 + \cos(\chi_f t) + \cos(2\chi_f t)$ with frequency content at $0, -2\chi_f, -4\chi_f$ has frequency content at $0, -\chi_f, -2\chi_f, -3\chi_f, -4\chi_f$ once it is squared $I(t)^2 = \frac{1}{2}(\cos(4\chi_f t) + 2\cos(3\chi_f t) + 5\cos(2\chi_f t) + 6\cos(\chi_f t) + 4)$. That is, selective pulses with monochromatic frequency content operate in the expected way, but frequency combs require additional engineering via optimal control or other manual means.

There are two primary benefits of the $gf/2$ scheme. The first is reduced leakage without choosing $\Delta \gg \alpha$, which substantially slows down gates. The frequency content of a pulse can cover up to $\alpha/2$ without colliding with a single photon transition. The second is that the driven Hamiltonian can be written in a rotating frame that removes fast oscillating terms modulating the envelope $\epsilon(t)$. This is accomplished simply by choosing $\Delta = \alpha/2$ in Eq. 7.1. Then the full transmon-cavity Hamiltonian is

$$\begin{aligned} H_{gf} =& \frac{\alpha}{2}A^\dagger A - \frac{\alpha}{2}A^\dagger A^\dagger A A - \frac{K_C}{2}C^\dagger C^\dagger C C - \chi_e|e\rangle\langle e|C^\dagger C - \chi_f|f\rangle\langle f|C^\dagger C \\ &+ \epsilon_A^*(t)A + \epsilon_A(t)A^\dagger + \epsilon_C^*(t)C + \epsilon_C(t)C^\dagger. \end{aligned} \tag{7.7}$$

When playing $gf/2$ control waveforms, we set the generator frequency to $\omega_A - \frac{\alpha}{2}$ to realize the same rotating frame as used above.

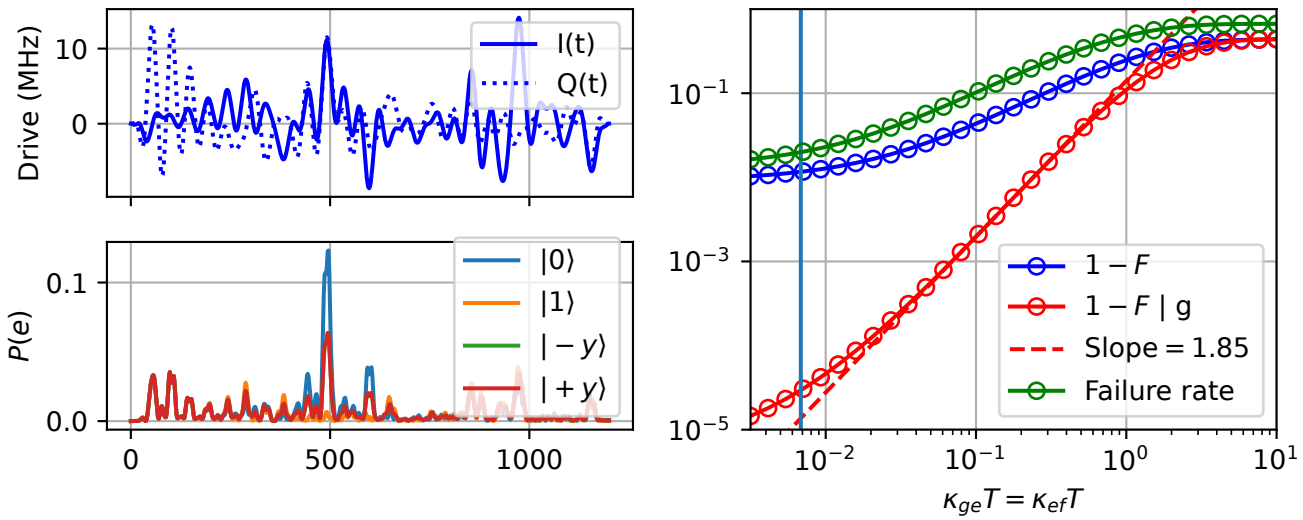


Figure 7.3: This figure has exactly the same format as Figs. 6.1 and 6.2. **Top left**: real $I(t)$ and imaginary $Q(t)$ parts of the control waveform $\epsilon(t)$ in Hamiltonian 7.7 that implements $R(Z = -\pi/2)$ on the binomial code with SNAP. The Hamiltonian parameters for the optimization are listed in Table 7.1. **Bottom left**: the probability of finding the transmon in $|e\rangle$ for four different initial states. **Right**: simulation of the resulting gate fidelities using Eq. 2.58 in the presence of only decay decoherence $\mathcal{D}[\sqrt{\kappa}|g\rangle\langle e|]+\mathcal{D}[\sqrt{\kappa}|e\rangle\langle f|]$. We set the decay rates between all three levels to be equal for simplicity. Accurate error detection of single decay then manifests with $O((\kappa T)^2)$ scaling. The exponent achieved here of 1.85 is likely limited by residual $|e\rangle$ populations during the pulse. The instantaneous leakage can be reduced with longer pulses, at the cost of a higher probability of any jump occurring.

Even though our choice of $\Delta = \alpha/2$ naturally reduces leakage to $|e\rangle$, there can still be residual leakage, especially as we try to minimize the length of our control pulses to reduce the effects of decoherence. We can ask optimal control to minimize the leakage by adding a cost function that explicitly computes the population of $|e\rangle$, but there is a more convenient option. The jump operator $\mathcal{D}[\sqrt{\kappa_{ge}}|g\rangle\langle e|]$ contributes a no-jump term to the effective Hamiltonian $-\frac{i\kappa_{ge}}{2}|e\rangle\langle e|$. The no-jump evolution U_{eff} reduces the norm of states based on the probability that no jump occurs during their trajectory. This offers a way to penalize intermediate population in $|e\rangle$: the no-jump evolution has no effect if the state's trajectory never contains $|e\rangle$. Furthermore, the optimal fidelity is reached when both the state transfers are optimized and the norm reduction from this no-jump term is minimized. This produces an optimization Hamiltonian

$$H_{\text{opt}} = H_{gf} - i\kappa_e|e\rangle\langle e| \tag{7.8}$$

where an effective choice of κ_e is typically on the order of $1/(500\text{ ns})$ for the optimizations done here. Choosing a κ_e that is too small blunts the effectiveness of the penalty, while choosing one too large overwhelms the optimal fidelity locations in the optimization landscape. A logical R_Z^L gate of the type discussed in Section 6.2.4 optimized using these leakage minimization techniques is shown in Fig. 7.3. Since the objective of the optimizer no longer simply describes the fidelity, but the fidelity with a penalty, the true fidelity of the gate needs to be simulated post-optimization. To test the quality of the leakage minimization, it's best to perform simulations such as the one in Fig. 7.3 and look for evidence of quadratic scaling in the presence of error detection. Poor leakage minimization will result in gates with linear scaling in $\kappa_{ge}T$.

Table 7.1: Hamiltonian parameters for Eq. 7.7 and system decoherence rates. The transmon gf dephasing rate Γ_f^ϕ varies rather widely in time due to a suspected coupling to a two-level system.

System Parameter	Operator convention	Value
Transmon frequency	$\omega_A A^\dagger A$	$\omega_A/2\pi = 4.038$ GHz
Transmon anharmonicity	$-\frac{\alpha}{2}A^\dagger A^\dagger AA$	$\alpha/2\pi = 134$ MHz
Cavity frequency	$\omega_C C^\dagger C$	$\omega_C/2\pi = 4.038$ GHz
Readout frequency	$\omega_r r^\dagger r$	$\omega_r/2\pi = 9.215$ GHz
Dispersive shift from $\|e\rangle$	$-\chi_e\|e\rangle\langle e\|C^\dagger C$	$\chi_e/2\pi = 386.4 \pm 0.4$ kHz
Dispersive shift from $\|f\rangle$	$-\chi_f\|f\rangle\langle f\|C^\dagger C$	$\chi_f/2\pi = 579.3 \pm 0.9$ kHz
Second order dispersive shift	$\frac{\chi_e'}{2}\|e\rangle\langle e\|C^\dagger C^\dagger CC$	$\chi_e'/2\pi = 0.4 \pm 0.1$ kHz
Cavity self-Kerr	$-\frac{K_C}{2}C^\dagger C^\dagger CC$	$K_C/2\pi = 507$ Hz
Transmon $\|e\rangle$ lifetime	$\mathcal{D}[\sqrt{\kappa_{ge}}\|g\rangle\langle e\|]$	$1/\kappa_{ge} \approx 54$ μs
Transmon $\|e\rangle$ thermal population		$P(\|e\rangle) = 0.01$
Transmon $\|g\rangle \to \|e\rangle$ heating	$\mathcal{D}[\sqrt{\kappa_{eg}}\|e\rangle\langle g\|]$	$1/\kappa_{eg} > 4.5$ ms
Transmon $\|f\rangle$ lifetime	$\mathcal{D}[\sqrt{\kappa_{ef}}\|e\rangle\langle f\|]$	$1/\kappa_{ef} \approx 42$ μs
Transmon ge dephasing	$\mathcal{D}\left[\sqrt{2\Gamma_e^\phi}\|e\rangle\langle e\|\right]$	$1/\Gamma_e^\phi \approx 300$ μs
Transmon gf dephasing	$\mathcal{D}\left[\sqrt{2\Gamma_f^\phi}\|f\rangle\langle f\|\right]$	$1/\Gamma_f^\phi \approx 30 - 300$ μs
Cavity single photon decay	$\mathcal{D}[\sqrt{\kappa_C}C]$	$1/\kappa_C \approx 1$ ms

7.2 Benchmarking the $g - f$ qubit

Once we've optimized or designed a gate, we want to evaluate its performance in actual quantum hardware. However, the fidelity functions in Section 2.2.2 assume full access to the state at the output of $\mathcal{U}\rho$. If we can reconstruct the states at the output of a channel, as in Section 3.2.1, then we can evaluate the traces to find the average gate fidelity. This type of state reconstruction requires full state tomography, which is particularly expensive (number of measurements needed) for cavity oscillators with many levels. Gate set tomography and process tomography reconstruct the channel $\mathcal{U}$ itself, but these approaches require even more measurements than state tomography. Is there a simpler way to evaluate gate fidelity? Yes!

Randomized benchmarking is a technique that enables the extraction of qubit gate fidelities without performing any gate or state tomography. It is SPAM (state

preparation and measurement) independent, making it a favorite technique of experimentalists. The main idea, which we will discuss further below, is to start with a set of gates that generate all rotations among the six cardinal states of the Bloch sphere. If one chooses a sequence of gates that ideally perform the identity, then the fraction of repetitions that return to the starting state gives us an idea of the accuracy of the constituent gates. Repeating this process for different numbers of gates provides a sequence-length dependent rate at which the state returns to its starting point. This rate is related to the average gate fidelity of the aforementioned set of gates.

7.2.1 The randomized benchmarking toolbox

Formally, randomized benchmarking (RB) is a technique to convert any error channel into the depolarizing channel [Magesan et al., 2011, Horodecki et al., 1999, Emerson et al., 2005]. The probability of an idling state under a depolarizing channel to remain in its starting state is exponential in time. The rate of exponential decay λ is the strength of the depolarizing channel which is related to the average gate infidelity $1 - \text{avg}_k \bar{F}(\mathcal{U}_k, C_k)$ for unitary gates $C_k \in \mathcal{C}$. The $\bar{F}$ used here is the same as defined in Eq. 2.61. The set $\mathcal{C}$ is required to be the 24 Clifford gates (or gates whose products generate all Cliffords) to produce the "twirling" effect that converts any error channel into depolarizing noise. The full theory of randomized benchmarking is beyond the scope of this dissertation, but we review four versions of randomized benchmarking that we use in the following sections and chapter. This review is not intended to be exhaustive, and focuses on showing which quantities one learns from each method and the experimental requirements of the method.

Clifford/generator randomized benchmarking (CRB/GRB)

Using random unitaries to estimate the average gate fidelity [Emerson et al., 2005] inspired the first version of randomized benchmarking [Knill et al., 2008] implemented

on trapped ions. NMR [Ryan et al., 2009] and superconducting qubit implementations [Chow et al., 2009] soon followed, establishing randomized benchmarking as a platform-agnostic tool to quantify average gate fidelities. Later work formalized the Clifford group $\mathcal{C}$ (or a group G that generates the Clifford group) as the set of random unitaries that provide the most rigorous estimate of average gate fidelity [Magesan et al., 2011]. The steps to implement CRB/GRB are:

1. Initialize the qubit in one of its logical states, say $|0\rangle$.

2. Choose n random gates U_k from $\mathcal{C}$ (G) and compute the recovery gate $U_{n+1} = U_1^\dagger U_2^\dagger \cdots U_n^\dagger$ that ideally reverses the evolution of $U_n \cdots U_2 U_1$.

3. Apply $U_{n+1} U_n \cdots U_2 U_1$ to $|0\rangle$ and measure $|0\rangle\langle 0|$.

4. Repeat many times at different n to build up statistics of the probability $P_0(n)$ to return to $|0\rangle$ after n random gates (this is often called the "survival probability."

5. Fit $P_0(n) = A + B\lambda^n$. The average infidelity of each Clifford (generator) gate in $\mathcal{C}$ (G) is then $(1-\lambda)(d-1)/d$ for the Hilbert space dimension d ($d = 2$ for qubits).

The circuit for RB/GRB is for $1 \leq k \leq n$. Note that each Clifford gate C_k may be composed of multiple generator gates [Epstein et al., 2014], so the infidelity per gate when λ is extracted from Clifford RB is then $\lambda/\bar{n}$ where $\bar{n}$ is the average number of generators needed per Clifford. Likewise, we can estimate a Clifford fidelity from GRB by multiplying that λ by $\bar{n}$.

Leakage randomized benchmarking (LRB)

[Wood and Gambetta, 2018, Andrews et al., 2019, McKay et al., 2017, Chen et al., 2016]. In quantum error correcting and detecting codes, we define the codespace as a subspace of the entire Hilbert space in which we encode logical information. However,

coherent and incoherent errors occur that produce leakage out of the codespace into a recoverable (or detectable) error space as well as a potentially unrecoverable subspace. Leakage is not trace-preserving, and is thus not natively handled by CRB/GRB. We often want to quantify this rate of leakage out of the codespace and potential seepage back into the codespace. LRB enables us to find these rates as well as the average fidelity when the gate acts on a state initially in the codespace. This fidelity includes the effect of leakage, but ignores any evolution occurring in the leakage space (including seepage), as our information is likely already spoiled. This fidelity is experimentally achievable by detecting and removing leakage *before* each gate. Note that LRB can produces higher gate fidelity estimate than CRB/GRB when the main source of error is leakage.

The main procedural difference between LRB and CRB/GRB is the addition of measurements that can distinguish between the codespace and leakage space, not only determine if the state after a sequence of gates is the same as the initial state or not. To perform LRB as in [Wood and Gambetta, 2018], follow steps 1-4 of CRB/GRB. At the end of these steps, we have two probabilities available to us: the survival probability $P_0(n)$ and the probability of remaining in the codespace $P_{\text{code}}(n)$. Fit $P_{\text{code}}(n) = A + B\lambda_{\text{code}}^n$. The leakage rate per gate is $L_1 = (1-A)(1-\lambda_{\text{code}})$ and the seepage rate is $L_1 = A(1-\lambda_{\text{code}})$. Next, use λ_{code} to fit $P_0(n) = A_0 + B_0\lambda_{\text{code}}^n + C_0\lambda^n$ where $0 \leq A_0 \leq A, 0 \leq C_0 \leq 1, 0 \leq A_0 + B_0 + C_0 \leq 1$. The average gate fidelity $\bar{F}$ is then $\bar{F} = ((d-1)\lambda + 1 - L_1)/d$ where d is the dimension of the codespace. If the leakage rate is very slow $\lambda_{\text{code}} \ll \lambda, B \ll A$, then we cannot accurately separate errors within the codespace from leakage and seepage errors. In this case, it is best to use conventional CRB/GRB.

Interleaved randomized benchmarking (IRB)

[Magesan et al., 2012]. This version of RB allows one to extract the infidelity of a single gate U by interleaving that gate in a random sequence of Cliffords $C_k \in \mathcal{C}$ In IRB, the gates between which we interleave U must be Cliffords, otherwise the IRB protocol is similar to CRB:

1. Initialize the qubit in one of its logical states, say $|0\rangle$.

2. Choose n random gates C_k from $\mathcal{C}$ and compute the recovery gate $U_{n+1} = C_1^\dagger U^\dagger C_2^\dagger U^\dagger \cdots C_n^\dagger U^\dagger$ that ideally reverses the evolution of $C_1 U C_2 \cdots C_n U$.

3. Apply $U_{n+1} U U_n \cdots U U_2 U U_1$ to $|0\rangle$ and measure $|0\rangle\langle 0|$.

4. Repeat many times at different n to build up statistics of the survival probability $P_0(n)$.

5. Fit $P_0(n) = A + B\lambda_U^n$. 6. Perform conventional CRB and extract λ as above.

6. The average gate infidelity of U is then $(d-1)(1-\lambda_U/\lambda)/d$. IRB works best when the infidelity of U is a similar order of magnitude to the infidelity of each Clifford C_k. When the true infidelity of U is about an order of magnitude less than C_k, the infidelity estimate begins to deviate from its true value [Epstein et al., 2014].

Character randomized benchmarking (cRB)

[Helsen et al., 2019, Claes et al., 2021]. cRB allows randomized benchmarking of a subset of gates (Clifford or non-Clifford) that form a group G. This is useful when there isn't access to the full set of Clifford gates or the Cliffords have substantially lower fidelity than a gate we want to benchmark with IRB. Examples of gate groups include the group $\{\mathbf{I}, R_Z(\pi), R_Z(\pi/2), R_Z(-\pi/2)\}$ in which the rotations around the Z axis can in principle be replaced with rotations around any other fixed axis. The

benchmarking protocol utilizes multiple different initial states and expectation measurements that depend on the exact structure of the group. The protocol is as follows:

1. Initialize the qubit in a logical state ρ_b.
2. Choose n random gates U_k from G. Compile the first two gates $U_2 U_1 = U_1'$ into a single gate and compute the inverse gate $U_{n+1} = U_1^\dagger U_2^\dagger \cdots U_n^\dagger$.
3. Apply $U_{n+1} U_n \cdots U_3 U_1'$ to the initial state ρ_b and measure an observable M_b.
4. Repeat many times at different n to build up statistics of each observable $M_b(n)$.
5. Compute a weighted survival probability as described in Eq. 3 of [Claes et al., 2021] and fit to Eq. 4. Then use eq. 5 to estimate the average gate fidelity of G.

Again, the exact form of ρ_i, M_i, and the weighted probability function depend on the group G. We explicitly list each of these when we benchmark the R_Z group mentioned above in Chapter 8.

7.2.2 $g - e$ qubit RB results

We've previously mentioned that many other efforts in quantum computing utilize the $g - e$ transmon qubit as the building block for error-correcting codes. Before benchmarking the $g - f$ qubit, we want to set a performance baseline and benchmark our transmon's $g - e$ qubit. Our system parameters are listed in Table 7.1. We implement all $g - e$ rotations with Gaussian pulses as in Eq. 2.45 and apply the first order DRAG correction $Q(t) = \dot{I}(t)/\alpha$ discussed in Section 2.2.1 resulting in the waveform shown in Fig. 2.1.

The LRB protocol allows us to extract the leakage rate, seepage rate, fidelity within the computational space, and overall fidelity. The results are shown in Fig. 7.4, but we want to emphasize a few of the results here as well. First, the overall fidelity

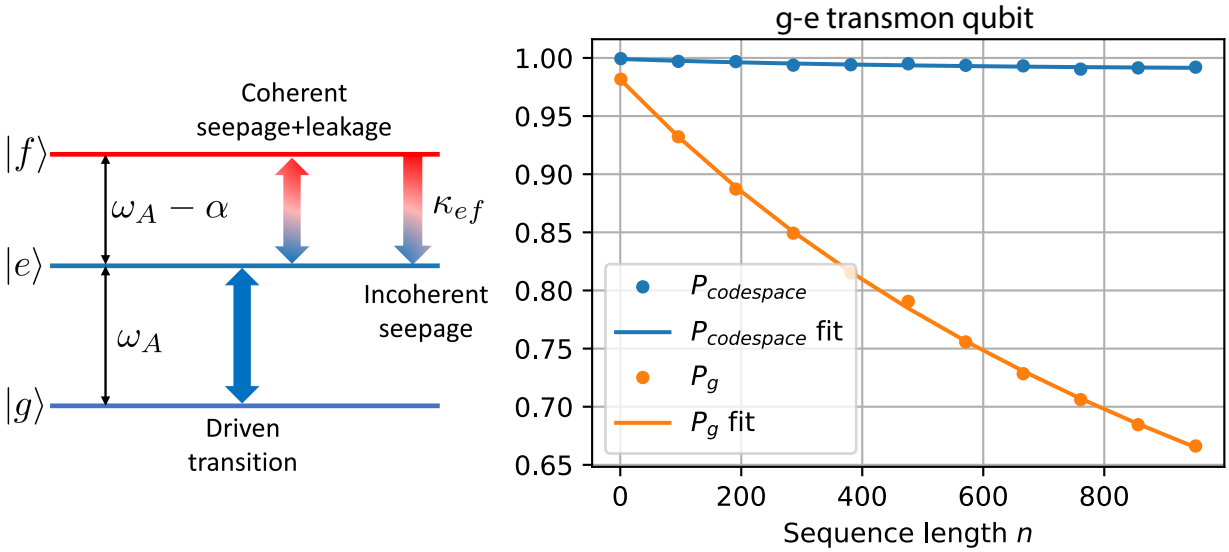


Figure 7.4: Leakage randomized benchmarking of the $g-e$ transmon qubit. **Left**: level diagram of the transmon qutrit. Our drive centered at ω_A primarily drives the $g-e$ transition, but also produces a small coupling between $e-f$. When the transmon leaks to $|f\rangle$ via this coupling, it randomly seeps back into the computational space via decay to $|e\rangle$ at rate κ_{eg}. **Right**: generator and leakage randomized benchmarking of the $g-e$ transmon qubit. Our generating set G here is $\{R_X(\pi/2), R_Y(\pi/2), R_X(\pi), R_Y(\pi)\}$ each with the waveform shown in Fig. 2.1, but with varying amplitude and phase. The GRB average fidelity is 0.99947 ± 0.00003. Our standard transmon readout techniques can distinguish the $|g\rangle, |e\rangle, |f\rangle$ states allowing us to perform LRB (higher transmon states are generally interpreted as $|f\rangle$ in our readout). The leakage rate is $L_1 = (2.0 \pm 1.2)10^{-5}$ and seepage rate is 0.002 ± 0.001 with average computational space fidelity 0.9998 ± 0.0004. The seepage rate is identical to the probability of a decay $|f\rangle \rightarrow |e\rangle$ during the 80 ns gate with $\kappa_{ef} = 2.5 \cdot 10^{-5}$/ns.

of the gate (which does not separate leakage and seepage into separate quantities) is 0.99947 ± 0.00003. We can achieve fidelities this high because the 80 ns gates are nearly three orders of magnitude less than any coherence time. Leakage to $|f\rangle$ is limited by the DRAG waveform correction to $(2.0 \pm 1.2)10^{-5}$ per gate. Seepage back into the computational space occurs with rate $(2 \pm 1)10^{-3}$ per gate which is approximately the gate time 80 ns divided by the $T_1^f = 40\mu$s.

7.2.3 $g-f$ qubit RB results

As mentioned in Section 7.1.2, we utilize optimal control to find $g-f$ qubit controls that minimize intermediate leakage to $|e\rangle$. We optimize a set of pulses $G = \{R_X(\pi/2), R_Y(\pi/2), R_X(\pi), R_Y(\pi)\}$ with a very aggressive duration of 80 ns in Fig. 7.5 using Eq. 7.7 with all cavity terms removed

$$H_{gf} = \frac{\alpha}{2}A^\dagger A - \frac{\alpha}{2}A^\dagger A^\dagger AA + \epsilon_A^*(t)A + \epsilon_A(t)A^\dagger \tag{7.9}$$

and include a penalty on $|e\rangle$ occupation Eq. 7.8. We use the unitary fidelity function Eq. 2.54 and limit the bandwidth of these control pulses to 50 MHz to avoid any non-idealities in upconverting these baseband pulses with IQ mixers.

When using numerically optimized waveforms, we must tune up a scale factor that converts between the unitless waveform amplitude and FPGA DAC amplitude. After roughly tuning this value, we perform fixed-depth randomized benchmarking while sweeping the scale factor over a small range $\pm 10\%$ to accurately tune this amplitude conversion factor. Randomized benchmarking can be used to tune up a variety of parameters [Rol et al., 2017, Reinhold, 2019], but we use it here only to tune this scale factor.

Now that we have an optimized set of $g-f$ qubit rotations, we perform LRB in Fig. 7.6 to test their performance. The average $g-f$ qubit gate fidelity reaches 0.998, short of the $g-e$ qubit fidelity with higher leakage and seepage rates. There are two main culprits: high leakage due to short 80 ns pulses and single excitation decay now manifests as leakage, where in the $g-e$ qubit it is a computational space error. Minimizing intermediate coherent leakage to $|e\rangle$ comes at the cost of higher exposure to incoherent excitation decay. We have already seen in Fig. 7.3 that low leakage rates that preserve transmon decay detection are possible for longer pulses. Erasure qubits based on the $g-f$ transmon qubit will prioritize coherent leakage to

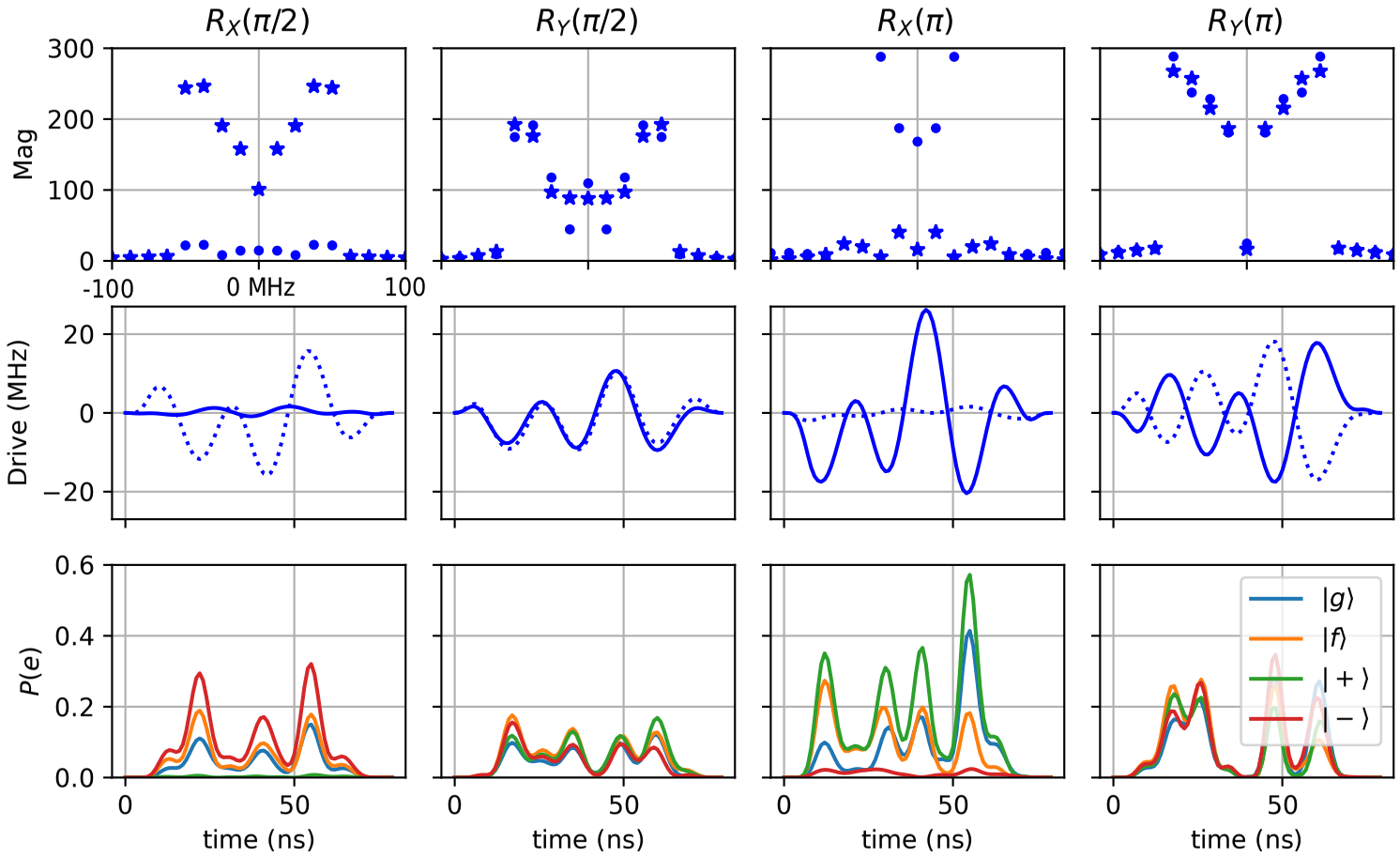


Figure 7.5: Waveforms and $|e\rangle$ population trajectories for our $g-f$ qubit generator set $G = \{R_X(\pi/2), R_Y(\pi/2), R_X(\pi), R_Y(\pi)\}$. The average gate fidelity Eq. 2.61 exceeds 0.999 in the absence of decoherence for each pulse. **Top row**: Magnitude of discrete Fourier transform components for $I(t)$ (dots) and $Q(t)$ (stars). All figures in the top row share the same x-axis. The pulses were optimized to cover 50 MHz of bandwidth, the sharp rolloff marks the 50 MHz boundary. **Middle row**: Time-domain waveforms of $I(t)$ (solid lines) and $Q(t)$ (dashed lines). These figures share a the time x-axis with the **bottom row**, which plots the $|e\rangle$ population trajectories for four different states listed in the legend. The $R_X(\pi)$ pulse has a surprisingly high amount of intermediate leakage to $|e\rangle$ compared to the other three pulses; we are not sure why the optimizer converged to this waveform.

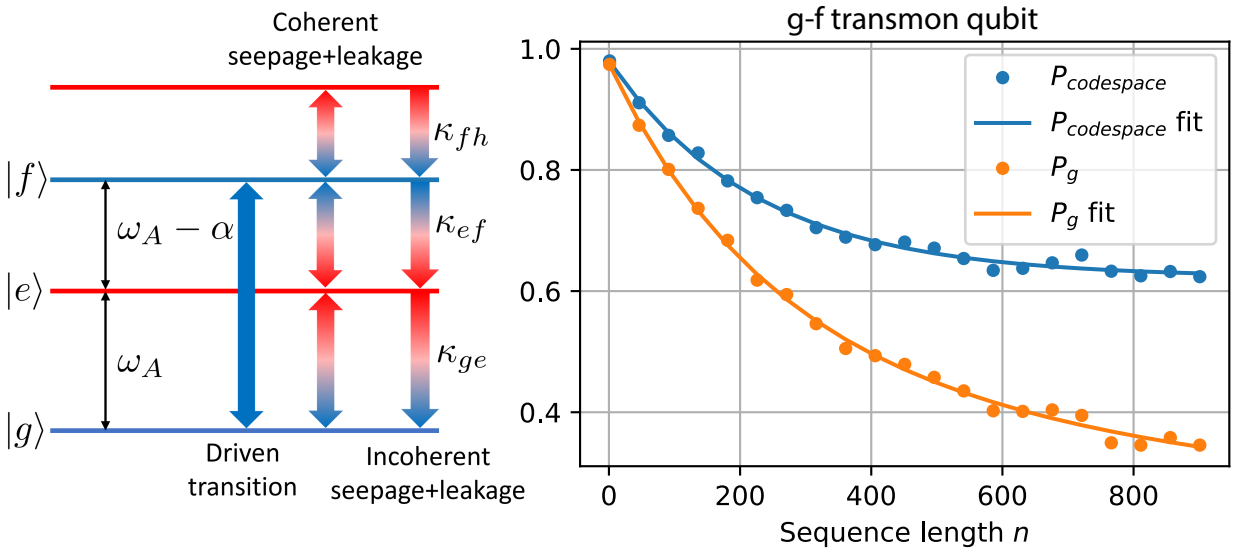


Figure 7.6: Leakage randomized benchmarking of the $g-f$ transmon qubit. **Left**: level diagram of the transmon qutrit. Our drive is centered at $\omega_A - \alpha/2$ to implement the $gf/2$ driving scheme. The strongest leakage channel is to $|e\rangle$, which can be reached by coherent residual couplings to $|e\rangle$ in Eq. 7.4 and $|f\rangle \to |e\rangle$ decays. **Right**: generator and leakage randomized benchmarking of the $g-f$ transmon qubit. Our generating set G here is $\{R_X(\pi/2), R_Y(\pi/2), R_X(\pi), R_Y(\pi)\}$ each optimized with Eq. 7.7 including the penalty Eq. 7.8 with waveforms shown in Fig. 7.5. The GRB average fidelity is 0.99840 ± 0.00007. The leakage rate is $L_1 = 0.00167 \pm 0.00008$ and seepage rate is 0.00275 ± 0.00013 with average computational space fidelity 0.9986 ± 0.0007. Like the $g-e$ pulses in Fig. 7.4, these control pulses have length 80 ns. The leakage rate is dominated by $|f\rangle \to |e\rangle$ whose probability is 0.002 as computed for the seepage case in Fig. 7.4. The discrepancy is likely due to the fact that $|f\rangle$ is not occupied with unity probability throughout the random gate sequence. The seepage rate exceeds the probability of $|e\rangle \to |g\rangle$ decay 0.0016 for $\kappa_{ef} = 2 \cdot 10^{-5}$/ns. This suggests that there is some residual coherent coupling to $|e\rangle$, which is a result of our choice to optimize as short of a pulse as possible. We can see intermediate leakage in the trajectories shown in Fig. 7.5. Increasing the pulse duration beyond 80 ns will reduce coupling to $|e\rangle$.

$|e\rangle$ that is not detectable [Kubica et al., 2022]. In Fig. C.1, we extend the pulse length to 160 ns which results in lower intermediate leakage. The relatively high fidelity of the $g-f$ qubit rotations without any error detection suggests that the SNAP gates discussed in Section 6.2.3, which only involve transmon $g-f$ qubit rotations, should achieve similarly high fidelities with error detection.

7.3 Adding a cavity drive

In this section, we extend $g-f$ control from just the transmon Hilbert space to the entire joint cavity-transmon Hilbert space. We provide two examples here that quantify the fidelity of binomial/kitten code state preparation and logical gates. As a reminder, the two logical binomial/kitten code words are

$$|0_L\rangle = |+Z\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |4\rangle) \tag{7.10}$$

$$|1_L\rangle = |-Z\rangle = |2\rangle\,. \tag{7.11}$$

State preparation pulses and logical gates are optimized with the $gf/2$ Hamiltonian in Eq. 7.7 with an added $|e\rangle$ population penalty and a penalty for states occupied above a cutoff

$$H_{\text{opt}} = H_{gf} - i\kappa_e|e\rangle\langle e| - i\kappa_N \sum_{n>N_{\text{max}}} |n\rangle\langle n| \tag{7.12}$$

where we generally choose $N_{\text{max}} = 10$ and $\kappa_N = 1/(100\text{ ns})$. The cutoff is important because we have to truncate the cavity Hilbert space in optimal control, and do not want spurious boundary effects resulting from population of the highest remaining levels.

7.3.1 Fock state creation

The most important binomial code word to initalize is $|1_L\rangle = |2\rangle$ since $|0_L\rangle$ can be prepared by applying a logical rotation to $|1_L\rangle$. We implement two photon state preparation by defining a unitary U_{prep} where $U_{\text{prep}}(|g\rangle \otimes |0\rangle) = |g\rangle \otimes |2\rangle$. Using the cost function 2.54 and Hamiltonian 7.12 with parameters in Table 7.1, we optimize a control pulse on the two dimensional subspace spanned by $|g\rangle \otimes |0\rangle\,, |g\rangle \otimes |2\rangle$. Since we don't define the action of U outside of these two states, the optimizer chooses a

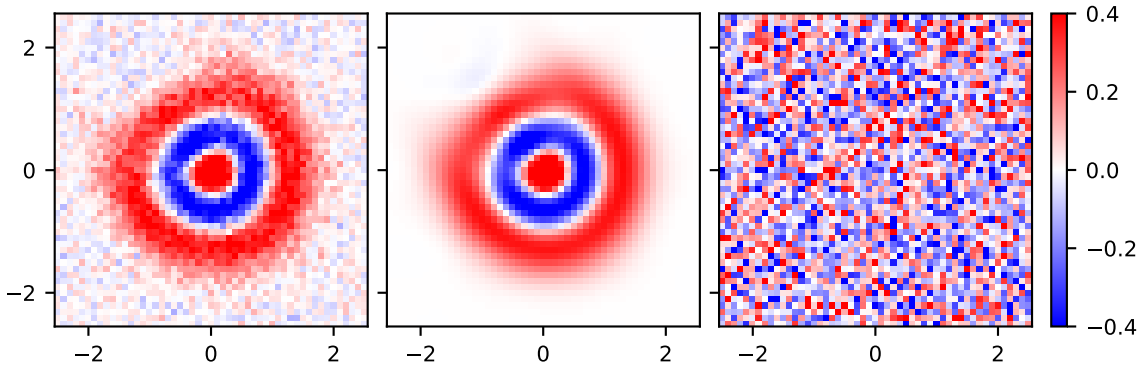
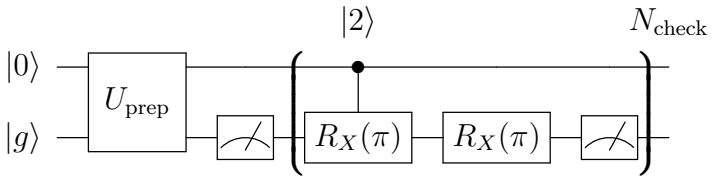


Figure 7.7: **Left**: measured Wigner function after apply a numerically optimized pulse to create $|2\rangle$ in the cavity from $|0\rangle$. The bandwidth of both the cavity and transmon control waveforms is restricted to 25 MHz. We reconstruct the density matrix from this Wigner using the techniques described in Section 3.2.1. **Middle**: plot of the Wigner function of the reconstructed state. The reconstructed state has fidelity 0.971 to the target state $|2\rangle$. **Right**: residuals between the measured and reconstructed Wigner functions, multiplied by a factor of 5. There is little structure in the residuals, indicating an accurate state reconstruction. The fidelity estimated here is limited by both state preparation and Wigner function measurement errors.

random, yet unitary, operation to perform on the rest of the Hilbert space. Since this gate is implemented with the $gf/2$ transition, we can detect when transmon decay occurs, but cannot always detect transmon dephasing. A Wigner function captured after preparing $|1_L\rangle$ is shown in Fig. 7.7. We reconstruct the density matrix from the Wigner function and estimate a fidelity of 0.971 between the prepared state and $|1_L\rangle$.

When preparing logical states, we always start by initializing the cavity and transmon in vacuum, apply the $|1_L\rangle$ pulse, and use selective π-pulses to verify preparation of two photons



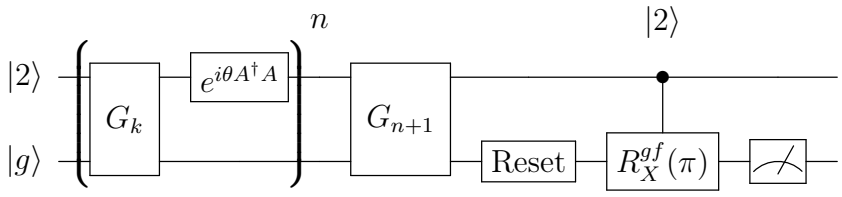
with $N_{\text{check}} = 2$ or 3. The transmon measurement after U_{prep} allows one to restart

state preparation if an error occurs U_{prep} that leaves the transmon in a superposition of ground and excited states. We only keep states for which all measurements return $|g\rangle$ which filters out many transmon errors during [Elder et al., 2020]. We can further increase the fidelity of our check by using a selective $g-f$ π-pulse instead of a $g-e$ one, which clearly flags transmon decay.

7.3.2 Binomial qubit RB

Now that we have logical state preparation, we can benchmark logical gates. Our goal in this section is to show the viability of using the $g-f$ transmon ancilla for gates on bosonic qubits. We return to error-detected gates in Chapter 8. We choose a generator set $G = \{R_X^L(\pi/2), R_Z^L(\pi/2), R_Z^L(-\pi/2), R_Z^L(\pi), \mathbf{I}\}$ that includes only one population-changing gate $R_X^L(\pi/2)$ that requires cavity drives. In all cases, we restrict the transmon drive bandwidth to 25 MHz and the cavity drive to 10 MHz and optimize with the trace fidelity Eq. 2.54 and Hamiltonian 7.7. We implement all $R_Z^L(\theta)$ gates with only drives on the transmon. Here, the identity gate is not instant, but takes the form of a 40 ns delay for technical reasons. Waveforms and leakage $|e\rangle$ population trajectories are shown in Fig. 7.9.

Unlike the transmon, we cannot in a single shot measure the binomial qubit into its code words or leakage states. For the moment, this rules out LRB, so we focus on GRB. Our benchmarking circuit is



where state preparation is performed as discussed in the previous section. Note that we initialize in $|1_L\rangle$ instead of $|0_L\rangle$ because $|1_L\rangle$ is both easier to prepare and

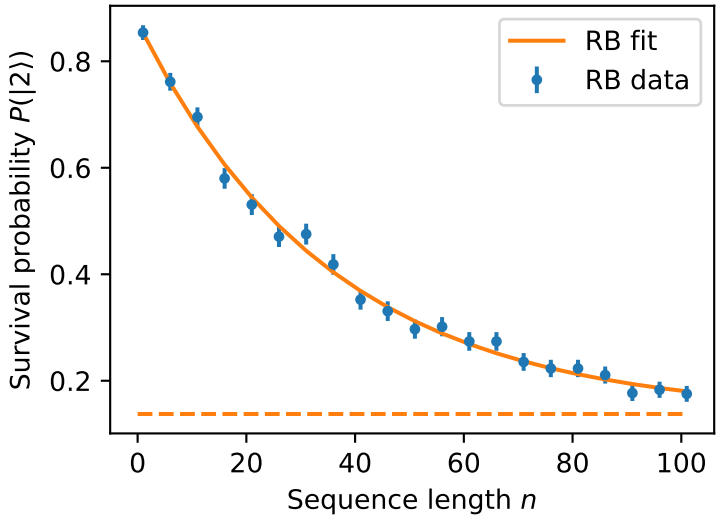


Figure 7.8: Randomized benchmarking of $G = \{R_X^L(\pi/2), R_Z^L(\pi/2), R_Z^L(-\pi/2), R_Z^L(\pi), \mathbf{I}\}$. All gates are generated via optimization of Eq. 2.54 with H_{gf} Eq. 7.12. Hamiltonian parameters are given in Table 7.1 and waveforms in Fig. 7.9. Each point in the figure is an average of 680 different random sequences of length n. The errorbars are the binomial uncertainty of finding the logical qubit in $|1_L\rangle$ at the end of the sequence. These errors are propagated through the fit to find an average gate fidelity of 0.986 ± 0.001. The dotted line represents the offset of the exponential fit, whose difference from 0.5 indicates the presence of leakage out of the codespace resulting from both transmon and cavity errors.

and measure. The cavity frame rotation $e^{i\theta A^\dagger A}$ with $\theta = -0.019\pi$ accounts for a miscalibration of χ_f and is not present in later experiments. The selective $g - f$ π-pulse at the end leaved the transmon in $|g\rangle$ only if there are not two photons in the cavity. Any measurements of $|e\rangle$ or $|f\rangle$ indicate two photons, where measurements of $|e\rangle$ indicate decay $|f\rangle \rightarrow |e\rangle$ during the selective pulse. This still indicates two photons in the cavity, as the transmon only reaches $|f\rangle$ when there are two photons.

Fig. 7.8 shows the results of this GRB circuit. The average gate fidelity of G reaches 0.986 ± 0.001, and is 0.983 ± 0.001 if we remove the near-unity fidelity identity gate from the average. These fidelities are comparable to numerically optimized gates with the $g-e$ transmon qubit [Heeres et al., 2017,Reinhold et al., 2020], demonstrating that the $g - f$ transmon is a drop-in replacement for the $g - e$ transmon.

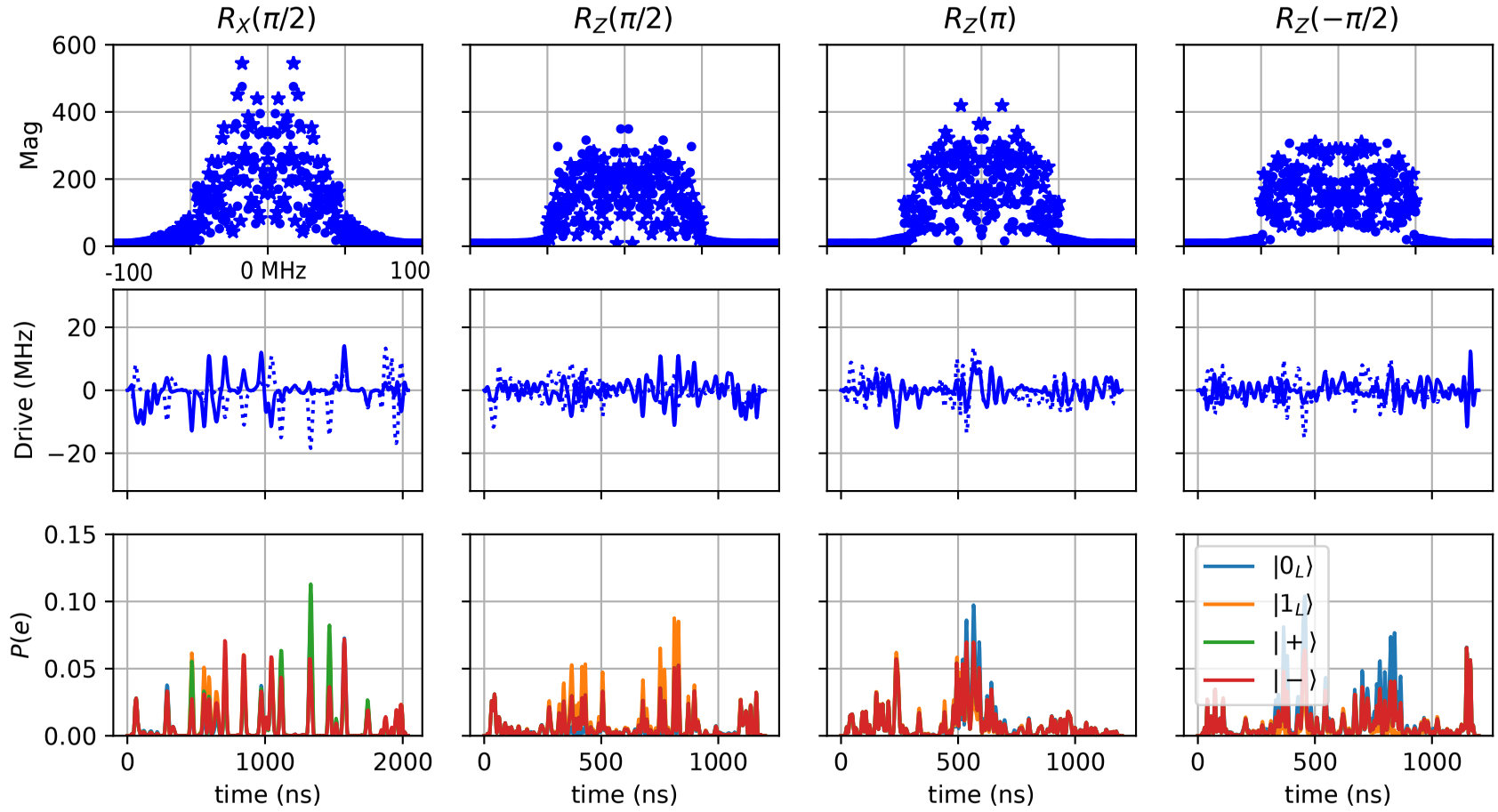


Figure 7.9: Transmon drive waveforms and $|e\rangle$ population trajectories for our $g-f$ qubit generator set $G = \{R_X^L(\pi/2), R_Z^L(\pi/2), R_Z^L(-\pi/2), R_Z^L(\pi), \mathbf{I}\}$. The average gate fidelity Eq. 2.61 exceeds 0.999 in the absence of decoherence for each pulse, except for the $R_X^L(\pi/2)$ pulse, which achieves a fidelity of 0.997. **Top row**: Magnitude of discrete Fourier transform components for the transmon controls $I(t)$ (dots) and $Q(t)$ (stars). All figures in the top row share the same x-axis. The cavity drive waveform for the $R_X(\pi/2)$ gate is not shown. The pulses were optimized to cover 50 MHz of bandwidth, the sharp roll off marks the 50 MHz boundary. **Middle row**: Transmon time-domain control waveforms of $I(t)$ (solid lines) and $Q(t)$ (dashed lines). These figures share a the time x-axis with the **bottom row**, which plots the $|e\rangle$ population trajectories for four different states listed in the legend. Notice that the duration of the $R_X^L(\pi/2)$ is longer than the Z rotations.

7.4 Conclusion

Our goal is the design and implementation of error-detected gates. In Chapter 6, we determined that we need a direct coupling to the third transmon level $|f\rangle$ so that any measurement indicating the transmon in $|e\rangle$ flags a transmon decay error. In this chapter, we utilized the $gf/2$ virtual two photon transition to engineer a direct $g-f$ interaction that enables fast $g-f$ qubit gates and is more amenable to optimal control techniques than previous Raman methods. Leakage to $|e\rangle$ limits the speed of the $g-f$ qubit gates, but can be mitigated with DRAG-like techniques and optimal control. We combine the $gf/2$ interaction with cavity drives to optimize binomial code logical state preparation and gates. Randomized benchmarking shows that these gates achieve the same fidelity as gates implemented with the $g-e$ transmon. In fact, we expect that in most cases the fidelity of gates implemented with the $g-f$ transmon will exceed the $g-e$ case, due to $\chi_f > \chi_e$, which is the relevant speed limit when combining cavity and transmon control.

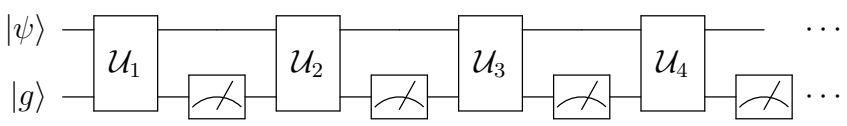
Chapter 8

Demonstration of error-detected gates

In the previous three chapters, we have laid out a set of requirements for an error-detected logical gate, introduced solutions to meet each requirement, and even demonstrated a $g - f$ qubit. Before moving on to our demonstration of an error-detected logical Z^L gate for the binomial code, we recap the requirements and our progress so far.

8.1 Recap

- **Principle of error-detected gates.** In Chapter, 5 we introduced the idea of adding a mid-circuit measurement after each gate that checks for detectable errors. A circuit



is only accepted if all measurements return successful results. The gain from error-detection is determined by two factors: our ability to map errors to a detectable state and the accuracy with which we can detect those states. Er-

ror detection improves the fidelity of gates marked successful by a factor of $p_{\text{phys}}/\epsilon_{\text{pass}}$ as discussed in Section 5.3. We showed that ϵ_{pass}, the rate at which we accept failed gates, is small due to high quality, QND transmon measurements which were characterized in Section 5.4. To realize the gain from error detection, we need to design gates that map errors to a specific set of detectable states.

- **Detection of transmon dephasing.** Jumps in the transmon are the primary source of errors for gates that require entangling the transmon and cavity. To build an error-detected gate, we need to design a gate that detects dephasing jumps. In Section 6.2, we introduce the idea of path independence as a framework for detecting harmful dephasing jumps. We introduce SNAP as an implementation of an error-detected R_Z^L gate. Path-independent SNAP requires a condition we call "equal-latitude," (discussed in Section 6.2.2) which can be difficult to enforce given the off-resonant Hamiltonian terms. We develop a version of optimal control in Section 6.2.4 that largely recovers first order error detectability as shown in Fig. 6.2.

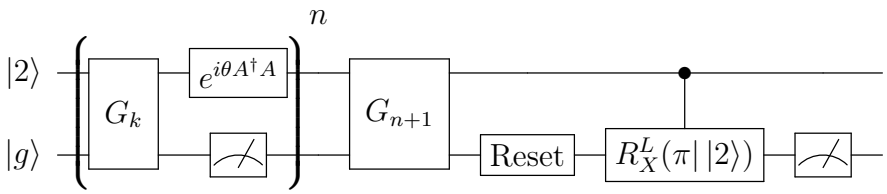
- **Detection of transmon decay.** The other main error mechanism is transmon excitation decay. In Section 6.2, we review the idea of using the lowest three levels of the transmon instead of the lowest two. If we can engineer a direct coupling between the $|g\rangle$ and $|f\rangle$ states, then any measurement of the transmon in $|e\rangle$ indicates a decay jump has occurred. Chapter 7 focuses on deriving and demonstrating this interaction. We utilize a virtual two photon transition to drive the $g-f$ interaction with only a single drive in Section 7.1.2. This makes numerically optimizing pulses easier, and enables $g-f$ qubit gates in Section 7.2.3 that approach the fidelity of $g-e$ qubit gates. The main limitation is set by our rather low transmon anharmonicity α. In Section 7.3, we expand

$g-f$ control by adding a cavity drive and optimize Fock state preparation as well as a set of logical binomial code gates with fidelity 0.983 ± 0.001. These results build our confidence that our $g-f$ control is flexible enough to enable high-quality error detection.

In this chapter, our goal is to characterize error-detected gates with randomized benchmarking. We introduce the benchmarking circuit in the next section, but observe that our $R_X^L(\theta), R_Y^L(\theta)$ rotations are of insufficient fidelity to accurately measure the $R_Z^L(\theta)$ gate fidelities with IRB as done in related implementations [Reinhold et al., 2020]. We instead switch to character randomized benchmarking (cRB) as defined in Section 7.2.1 to measure the average fidelity of our Z^L gates. The average gate infidelity extracted from cRB decreases by a factor of 22 when we include error correction compared to the case without. This improvement is powered by a somewhat surprising property of our gates.

8.2 Error-detected generator randomized benchmarking

As mentioned above, we did not use IRB to measure Z^L gate fidelity, but we will still include results from the RB results that establish the baseline for IRB. We start by modifying the GRB circuit used in Section 7.3.2 by adding a mid-circuit measurement



after each gate in the generator set $G = \{R_X^L(\pi/2), R_Z^L(\pi/2), R_Z^L(-\pi/2), R_Z^L(\pi), \mathbf{I}\}$. As discussed in Section 5.4, there is a small cross-Kerr between the readout resonator

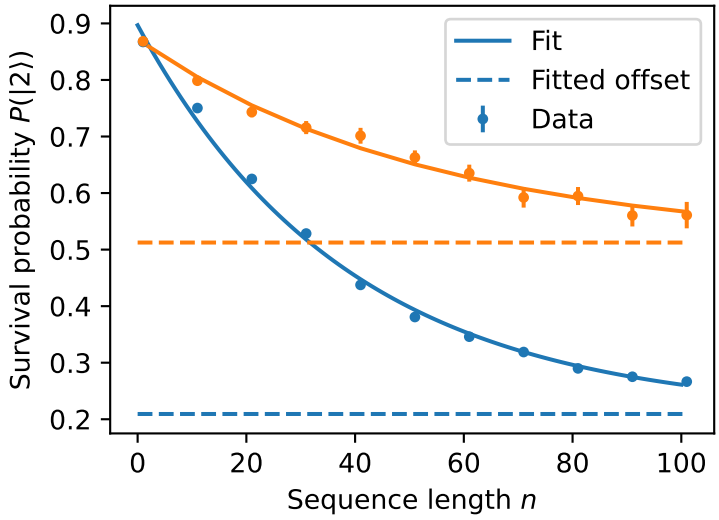


Figure 8.1: Randomized benchmarking of $G = \{R_X^L(\pi/2), R_Z^L(\pi/2), R_Z^L(-\pi/2), R_Z^L(\pi), \mathbf{I}\}$. All gates are generated via optimization of Eq. 2.54 from H_{gf} Eq. 7.12. This means that the R_Z^L gates here are not optimized for path-independence to dephasing, which we do not expect to have a significant impact at native coherence times. Hamiltonian parameters are given in Table 7.1 and waveforms in Fig. 7.9. The errorbars are the binomial uncertainty of finding the logical qubit in $|1_L\rangle$ at the end of the sequence. When we ignore the measurement results (except the very last measurement before decoding, blue line), we find an average gate fidelity 0.9872 ± 0.0003. Post-selecting on shots where every single measurement indicated $|g\rangle$ (orange line), we find an error-detected average gate fidelity of 0.991 ± 0.001. The dotted line represents the offset of each exponential fit, whose difference from 0.5 indicates the presence of leakage out of the codespace resulting from both transmon and cavity errors.

and the cavity, which imparts a small, deterministic phase to the cavity. We calibrate the deterministic phase $\theta = 0.034\pi$ phase per gate using the same Ramsey technique as in Fig. 5.4. For the results depicted in Fig. 8.1, the transmon reset is replaced with a simple measurement. At the time the data was taken, fast, repetitive measurements activated a process that excited the transmon into states higher than $|f\rangle$. The fidelity we extract from this circuit is not the average fidelity of the gates in G, but the average fidelity of these gates followed by a transmon measurement. If our transmon measurement is highly non-QND, we expect the non-error-detected results to be substantially worse than those in Fig. 7.8.

Results using the pulses optimized in Fig. 7.9 are displayed in Fig. 8.1. These pulses are not optimized to be path independent to transmon dephasing, so we only

expect to detect transmon decay in this experiment. We find an average gate fidelity of 0.9872 ± 0.0003 when ignoring the mid-circuit measurement results, which is comparable to the fidelity 0.986 ± 0.001 without any measurements discussed in Section 7.3.2.

Utilizing the mid-circuit measurement results to perform error-detection in Fig. 8.1 reduces the average gate infidelity from 0.0128 to 0.009, a reduction of only a factor of 1.42. We suspect that this rather disappointing result has two separate causes:

1. Anomalously high control infidelities of the $R_X^L(\pi/2)$ rotation. Other benchmarking data with a different generator set $G' = \{R_X^L(\pi/2), R_Y^L(\pi/2), R_X^L(\pi), R_Y^L(\pi)\}$ found non-error-detected average gate fidelities of ~ 0.97. The source of the infidelity is unknown, but distortion and delay of the control signals by the room temperature electronics is a prime candidate that can be mitigated with calibration. As the control infidelities increase, the gain from error-detection decreases because the infidelity is no longer dominated by incoherent jumps in the transmon and cavity. We are currently unaware of any path-independent implementations of $R_X^L(\theta)$ and $R_Y^L(\theta)$ rotations as the previously discussed exponential gadget construction is only compatible with excitation-conserving gates. The connection to the equal-latitude condition when the cavity is driven is also currently unknown.

2. No fast transmon reset. As mentioned above, the fast, repetitive readouts present in active transmon reset heat the transmon into a high excited state. In lieu of active reset, we post-select on the transmon being in the ground state following the gate G_{n+1} before the selective π-pulse checks if the cavity returned to its starting state. This post-selection inherently detects transmon errors that occur near the end of the random sequence even if we ignore the other mid-circuit measurements. The net result is that this approach *always* detects some transmon errors, and thus overestimates the baseline "non-error-detected" gate

fidelities. Later experiments in this chapter restore transmon reset and avoid this problem.

One surprising result in Fig. 8.1 is the reduction of leakage out of the binomial code space when error-detection is used. Survival probabilities below 0.5 indicate that leakage has occurred in addition to depolarization. The first occurrence of photon loss shifts code words to unique, orthogonal error words [Michael et al., 2016]

$$C\,|0_L\rangle = \sqrt{2}\,|\bar{0}_L\rangle = \sqrt{2}\,|3\rangle \tag{8.1}$$

$$C\,|1_L\rangle = \sqrt{2}\,|\bar{1}_L\rangle = \sqrt{2}\,|1\rangle\,. \tag{8.2}$$

That is, given a single photon jump during a string of gates in any RB sequence, the survival probability is zero. However, when we optimize our gates with the trace fidelity Eq. 2.54, we choose the projector P onto the binomial code space. In error detection, we do not need to preserve logical information in the error space, which lifts a constraint on gate design that results in substantially faster R_Z^L gates than in previous implementations [Reinhold et al., 2020], even with a smaller value of χ_f. Since we don't care about information encoded in the odd Fock states, the optimizer has the freedom to implement any unitary on those states. Sometimes this unitary leaves the transmon in a superposition of $|g\rangle$ and $|f\rangle$, adding some probabilistic cavity photon loss error detection since we flag measurements of $|f\rangle$ as an error. In Fig. 8.2, we plot probabilities that our gates excite the transmon to $|f\rangle$ given that the input to the gate is $|g\rangle \otimes |\bar{0}_L\rangle$ or $|g\rangle \otimes |\bar{1}_L\rangle$. We see that $\{R_X^L(\pi/2), R_Z^L(\pi/2), R_Z^L(-\pi/2)\}$ detect some error words with probability 0.5, while the $R_Z^L(\pi)$ gate does not detect any photon loss. This effect conveniently brings partial cavity photon loss detection without needing to add any additional parity measurements.

We will show shortly in Section 8.4 that the fidelity of all of the $R_Z^L(\theta)$ gates significantly exceeds that of the $R_X^L(\pi/2)$ gate. The R_Z^L are predicted to achieve non-

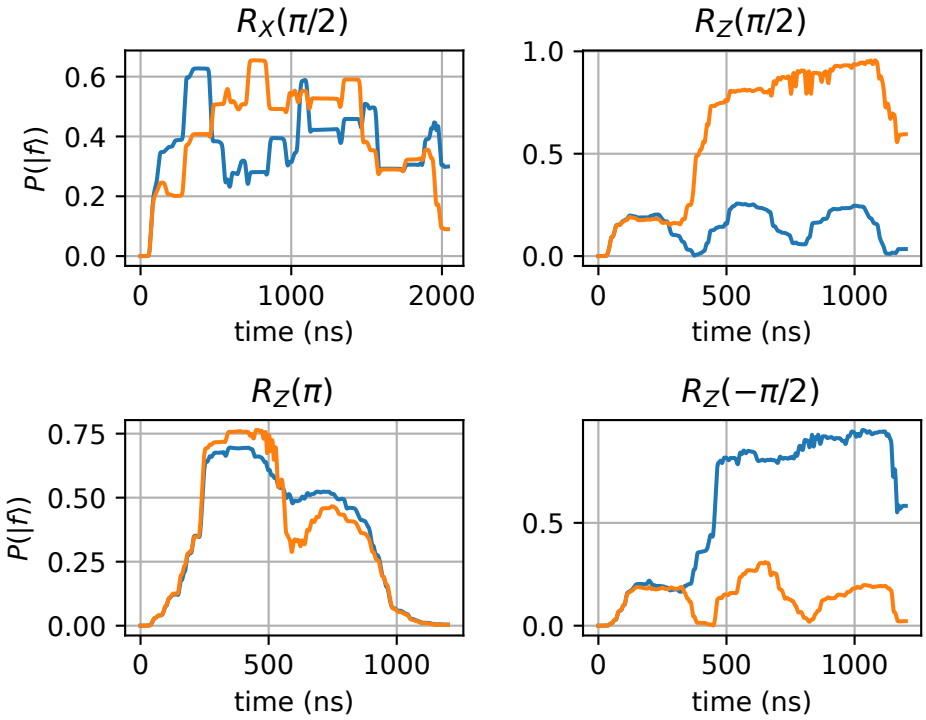


Figure 8.2: Each gate in the generator set $G = \{R_X^L(\pi/2), R_Z^L(\pi/2), R_Z^L(-\pi/2), R_Z^L(\pi), \mathbf{I}\}$ responds differently when the cavity input states are the error words $|g\rangle \otimes |\bar{1}_L\rangle$ (blue lines) or $|g\rangle \otimes |\bar{0}_L\rangle$ (orange lines). Three of the gates partially excite the transmon, so a subsequent transmon measurement of $|f\rangle$ will flag a logical error in the absence of additional transmon errors. These trajectories are simulated without any decoherence.

Table 8.1: cRB experiment configurations. The table is sorted by the initial state prepared at the beginning of each cRB shot. Note that $\rho_{b=0}$ isn't a physical state, we prepare $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$ in separate shots and subtract the Z^L expectation values of each to include the minus sign in $\rho_{b=0}$.

b	ρ_b	E_b	$\chi_b((S^L)^\alpha)$	$F_b(n)$	
0	$\frac{1}{2}(\|0\rangle\langle 0\| - \|1\rangle\langle 1\|)$	Z^L	1	$A_0\lambda_0^n + B_0$	$A_0 \approx 1, B_0 \approx 0$
1	$\|+\rangle\langle +\|$	X^L	$e^{i\pi\alpha/2}$	$A_1\lambda_1^n$	$A_1 \approx \frac{1}{2}$

error-detected fidelities approaching 0.99 (measured in Section 8.4), largely due to their shorter duration of 1200 ns vs. 2000 ns as seen in Fig. 7.9. In order to proceed to IRB, the infidelity of our Clifford gates must be less than ten times the infidelity of the interleaved gate. We need on average 2.25 elements of G to generate a Clifford, implying an error-detected Clifford infidelity of $(1-0.991)\cdot 2.25 = 0.020$. Simulations suggest that error-detected R_Z^L gates may reach or exceed fidelities of 0.998, so we abandon IRB and switch to cRB, which allows us to benchmark groups consisting solely of the high-fidelity R_Z^L gates.

8.3 Character randomized benchmarking protocol

We repeat the cRB steps here, this time tailored to the group $G_S = \{\mathbf{I}, R_Z^L(\pi), S^L, (S^L)^\dagger\}$ for $S^L = R_Z^L(\pi/2)$ [Claes, 2023].

1. Initialize the qubit in a logical state ρ_b from Table 8.1.

2. Choose n random gates U_k from G_S. Compile the first two gates $U_2U_1 = U_1'$ into a single gate and compute the inverse gate $U_{n+1} = U_1^\dagger U_2^\dagger \cdots U_n^\dagger$.

3. Apply $U_{n+1}U_n \cdots U_3U_1'$ to the initial state ρ_b and measure the observable E_b in Table 8.1.

4. Repeat many times at different n to build up statistics of each observable $\langle E_b\rangle(U_1, n)$. Note that the expectation value is a function of the circuit depth

n and initial gate U_1.

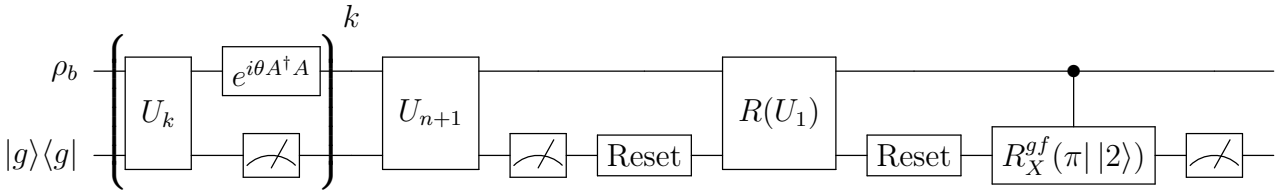
5. Compute the weighted survival probabilities $F_b(n) = \frac{1}{4}\sum_{U\in G_S}\chi_b^*(U)\,\langle E_b\rangle\,(U_1, n)$.

6. Fit each $F_b(n)$ according to Table 8.1.

7. Compute the average gate fidelity of G_S as $F_G = (\lambda_0 + 2\Re(\lambda_1) + 3)/6$.

Unlike CRB and GRB, cRB utilizes Pauli expectation values instead of survival probabilities. This presents an experimental difficulty because measuring Pauli expectation values of the binomial code is non-trivial. The primary challenge is that to measure a single-shot of an expectation value, we need to distinguish three different subspaces in the Hilbert space: both code words and the leakage/error space, much like leakage randomized benchmarking. Unfortunately, it is very difficult to map a three-outcome operator to the transmon; it is much easier to map the outcomes of binary-valued observables to the transmon, as in Chapter 4.

The easiest, and most accurate, measurement we can make on the cavity is the one we utilized during our benchmarking in 7.3.2: number-selective π-pulses. We therefore modify the cRB protocol to utilize the survival probability $P(U_1, n)$ instead of logical expectation values. This modification does not change the decay rates of the fitted exponentials, and thus does not affect the resulting fidelity. Our circuit is then



where the axis of logical rotation $R^L(U_1)$ before the selective π-pulse is determined by U_1 in each shot to ideally bring the logical state back to $|1_L\rangle$ at the end of the sequence. If these logical rotations are not all of nearly equal fidelity, this addition

can spoil cRB's insensitivity to SPAM error. We use a number-selective $g-f$ π-pulse at the end that correctly indicates $|2\rangle$ even in the presence of transmon decay, as any path to $|e\rangle$ must first pass through $|f\rangle$ (assuming that heating is negligible). This technique substantially improves measurement contrast.

8.4 Character randomized benchmarking results

8.4.1 Gate optimization

At this point, we want to put everything together and optimize waveforms for the set G_S that utilize the $gf/2$ transition and are path independent to one dephasing jump. Before optimizing, we observe that there is a 2876 ns delay between gates in cRB due to the readout duration and phase update latency. During this idle time, cavity Kerr is still present, but can be preemptively cancelled by adjusting the phases of the SNAP gates implementing each element of G_S because cavity Kerr commutes with the $gf/2$ SNAP Hamiltonian

$$\begin{aligned} H =& \frac{\alpha}{2}A^\dagger A - \frac{\alpha}{2}A^\dagger A^\dagger AA - \frac{K_C}{2}C^\dagger C^\dagger CC - \chi_e|e\rangle\langle e|C^\dagger C - \chi_f|f\rangle\langle f|C^\dagger C \\ &+ \epsilon_A^*(t)A + \epsilon_A(t)A^\dagger. \end{aligned} \tag{8.3}$$

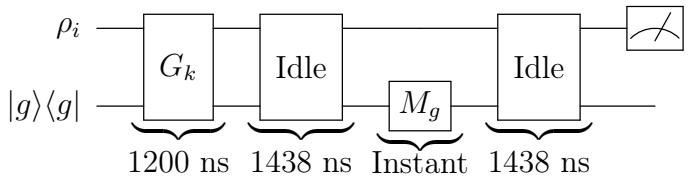
For a similar reason, rather than implementing the identity gate $\mathbf{I}$ as a brief delay followed by a readout, which must be present after each gate for technical reasons (the number of readouts per experimental shot must be constant in our control system), we choose to optimize a pulse with the same length as the R_Z^L gates. All pulses are optimized using the path independent cost function in Eq. 6.53 with $U_t = R_Z^L(\theta)$ defined only on the logical code space formed from even Fock states. We include a code snippet used to generate the pulses in Appendix D. We plot waveforms and latitudes in Fig. C.2. In Fig. 8.3, we plot the trajectories for the error space $|\bar{0}_L\rangle$ and

$|\bar{1}_L\rangle$ as well as the infidelity scaling with respect to only transmon decay or dephasing. The near $O(p_{\text{phys}}^2)$ scaling to both decay and dephasing jumps suggests that our pulses should produce error-detected gates. We note that the identity pulse does not display $O(p_{\text{phys}}^2)$ scaling to dephasing, but is higher fidelity than the R_Z^L rotations because very little excited state population is required to implement the Kerr-corrected identity gate.

8.4.2 Fidelity and benchmarking simulations

While the infidelity plots in Fig. 8.3 suggest that our gates could be error detectable, we would like to predict their overall performance in the presence of all errors as well. In this section, we perform two simulations of the gate fidelities. The first computes the fidelity of each gate in G_S separately using Eq. 6.49 while the second simulates the cRB protocol itself to capture the effect of probabilisitic cavity error detection. We will see that the average gate fidelity extracted from cRB simulations exceeds the single-shot average gate fidelity because of this effect.

To simulate the fidelity of a single gate, we must simulate the gate evolution itself, as well as idle evolution during the readout. Since the readout takes twice the duration of the gate (2876 ns vs. 1200 ns), we expect the readout to introduce significant infidelity via cavity decay. To compute the average gate fidelity with the readout, we simulate the evolution of four states $\rho_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |0\rangle\langle 0|, |1\rangle\langle 1|\}$ under the following circuit



that models the transmon readout as instant projection sandwiched between idle evolution [Chou et al., 2023]. M_g is the measurement operator corresponding to the

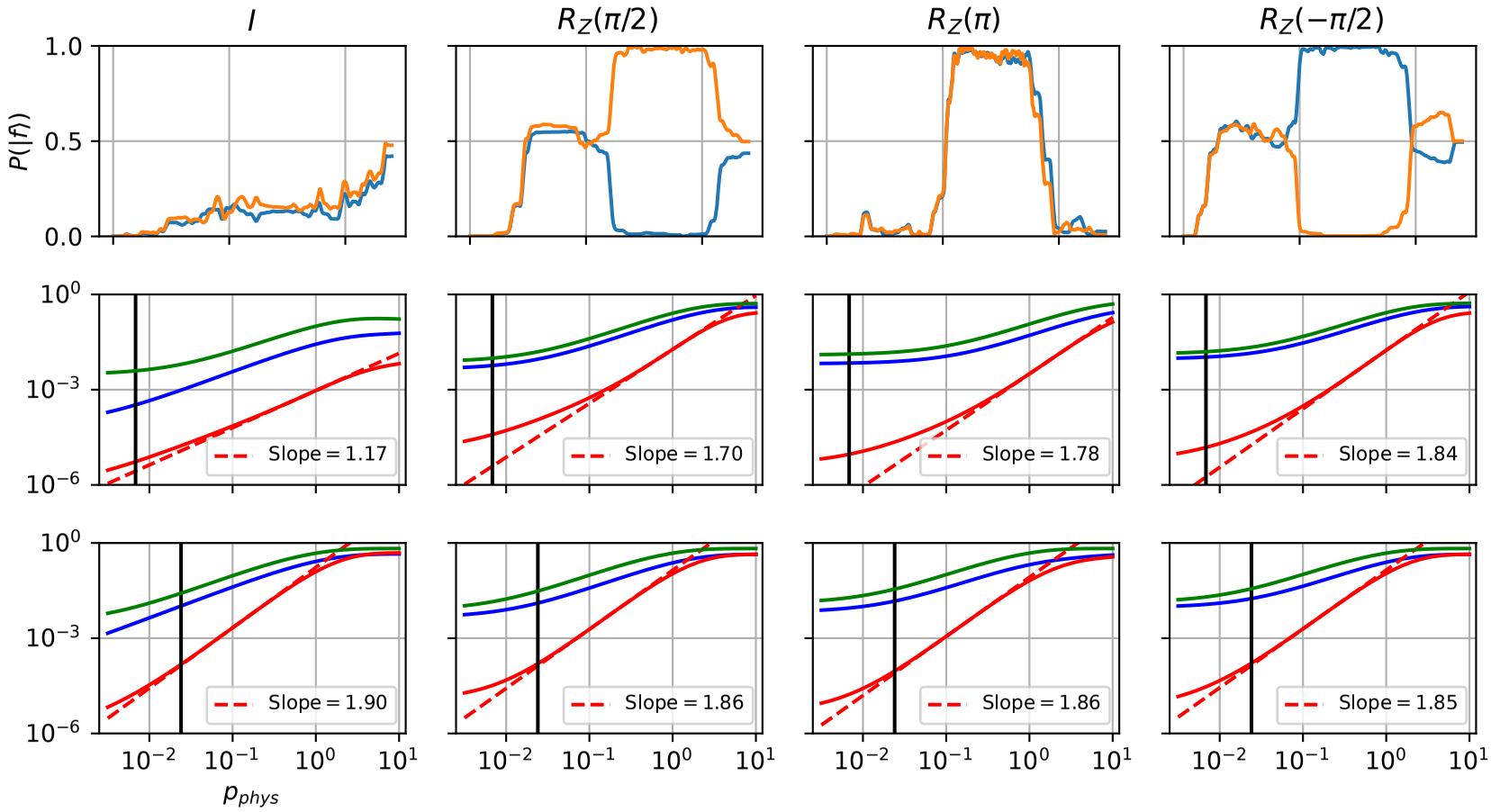
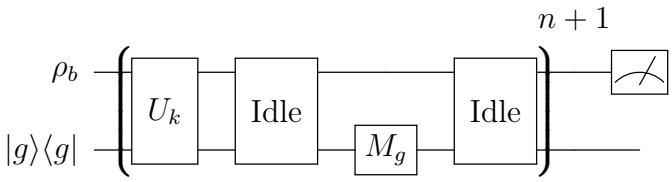


Figure 8.3: Error scaling for the pulse displayed in Fig. C.2. **Top row**: probability of finding the transmon in $|f\rangle$ after each gate when the input state is $|g\rangle \otimes |\bar{1}_L\rangle$ (blue) or $|g\rangle \otimes |\bar{0}_L\rangle$ (orange). The time x axis is not displayed, but ranges up to the 1200 ns length of the pulses. **Middle row**: error scaling with only transmon dephasing noise. The green line is the success probability, blue is the non-error detected fidelity, and red is the fidelity with perfect error detection. The dotted line is the tangent line with maximum slope marking where the highest order of scaling in p_{phys} occurs. These values are computed with Eq. 2.58 in the presence of only transmon dephasing $\mathcal{D}\left[\sqrt{2\Gamma_f^\phi}|f\rangle\langle f|\right]$. The x axis is the unit-less single jump probability $p_{\text{phys}} = \Gamma_f^\phi T$ where $T = 1200$ ns is the pulse duration. The vertical black line marks $1/\Gamma_f^\phi = 176$ μs which is a common value for the system used in this experiment. **Bottom row**: error scaling with both transmon decays $\mathcal{D}[\sqrt{\kappa_{ge}}|g\rangle\langle e|] + \mathcal{D}[\sqrt{\kappa_{ef}}|e\rangle\langle f|]$. For simplicity, we scale the jump rates together $\kappa = \kappa_{ge} = \kappa_{ef}$ [Tsunoda et al., 2023] so that the x axis is $p_{\text{phys}} = \kappa T$. The vertical black lines mark $1/\kappa = 50$ μs, which is a common value in our experiment.

ground state $|g\rangle$. This is the point where we can add measurement infidelity and non-QNDness. The final cavity measurement marks where we compute the average gate fidelity. We plot results of this fidelity in Fig. 8.7 using Hamiltonian 8.3 with all decoherence operators characterized in Table 7.1. The result here is the fidelity averaged over all gates in G_S. The slope in the error scaling is depressed by a rather high baseline error from cavity photon loss during the gate. Our simulation suggests that decreasing the readout duration would substantially improve single-gate fidelities.

The second simulation mimics exactly the cRB protocol, while incorporating the technique used above to model the mid-circuit transmon measurement. The circuit is similar starting with a cRB initial state ρ_b



with the same durations as used in the average gate fidelity simulation above. Here, instead of simulating a new state trajectory for every random gate sequence, we compute the superoperator propagators corresponding to each block in the circuit diagram and use them to propagate the state through the random sequence of G_k. After each application of M_g, we re-normalize the state and compute the relevant logical expectation value. With these expectation values, we can extract an average gate fidelity for G_S by fitting the weighted survival probabilities. Results of these simulations at the experimentally used gate depths are plotted in Figs. 8.4 and 8.7.

8.4.3 Experimental results and discussion

To implement the protocol in Section 8.3, we must prepare three different states for the $b = 0$ and $b = 1$ components of cRB. For $b = 1$, we prepare $\rho_{b=1} = (|0_L\rangle + |1_L\rangle)/\sqrt{2}$ and apply random gates sorted by the first one before measuring in the logical X^L

basis. This component of cRB measures the accuracy of our logical rotations between superpositions of $|0_L\rangle, |1_L\rangle$. The $b = 0$ stage of cRB prepares eigenstates of G_S (also the code words) to measure the rate of bit flips and leakage induced by our gates. The initial state $\rho_{b=0} = (|0_L\rangle\langle 0_L| - |1_L\rangle\langle 1_L|)/2$ for this stage is traceless and thus unphysical. However, we achieve the same result by preparing each of $|0_L\rangle, |1_L\rangle$, applying random gates, and measuring the probability of remaining in the initial state. The minus sign in $\rho_{b=0}$ is added in post-processing to compute the expectation value $\langle Z^L \rangle$ for the initial state $\rho_{b=0} = (|0_L\rangle\langle 0_L| - |1_L\rangle\langle 1_L|)/2$.

In Fig. 8.4, we show cRB results for the $b = 0$ and $b = 1$ cases up to gate depth $n = 71$. For clarity, we do not combine the outcomes of each $b = 0$ state preparation into a single expectation value. Using the mid-circuit measurements to perform error detection clearly provides a substantial improvement in gate fidelity. The improvement is so significant that we can only see linear decay in the error detected survival probabilities. The measured success probability of 0.9634 ± 0.0003 limits the benchmarking sequence length and prevents us from seeing the decay plateau. At the longest sequence of 71 gates, the sequence completes with no detected errors only 7.4% of the time. There are a couple of potential problems that merit discussion.

The first is that in the $b = 0$ case, the survival probability of code words $|0_L\rangle$ and $|1_L\rangle$ have noticeably different decay rates with and without error detection. One source of this discrepancy is that the probabilistic cavity error detection provided by the gates (see Fig. 8.3) detects jumps in both code words, but not residual no-jump evolution in $|0_L\rangle$. The other source is the static cavity self-Kerr that is only preemptively cancelled when the transmon does not experience a jump. $|1_L\rangle$ is immune to transmon jumps of any type because they do not induce leakage, unlike the code word $|0_L\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |4\rangle)$ that can leak to $\frac{1}{\sqrt{2}}(|0\rangle - |4\rangle)$ due to transmon errors that spoil the superposition phase. This phase sensitivity is why the survival probabilities for $\rho_{b=0} = |0_L\rangle\langle 0_L|$ decrease so much faster than those for $\rho_{b=0} = |1_L\rangle\langle 1_L|$ in Fig. 8.4.

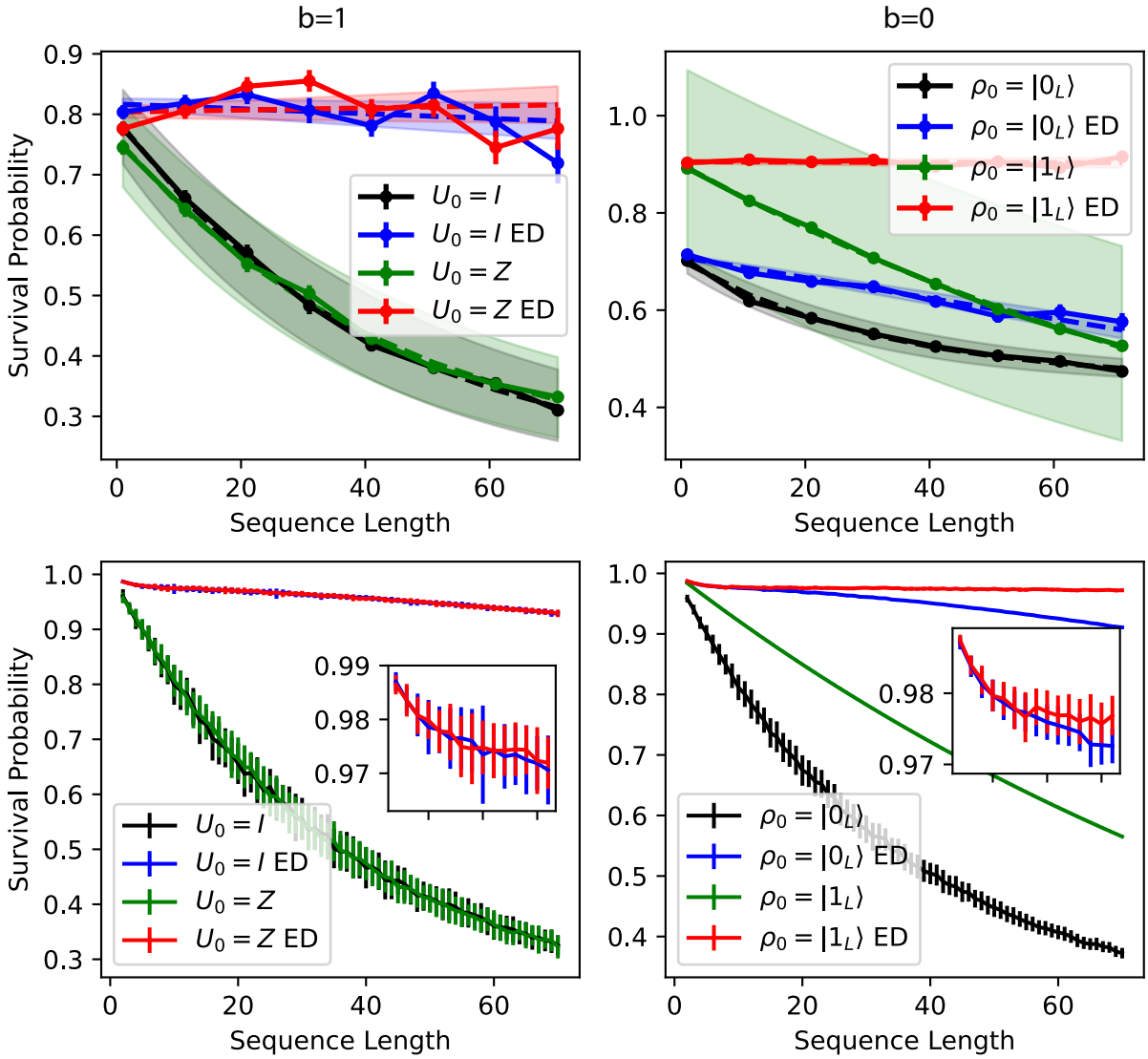


Figure 8.4: **Top row**: experimental cRB results. The two cases $b = 0$ and $b = 1$ are carried out according to the parameters in Table 8.1. The fit function for each data trace is linear $An + B$ or exponential $A\lambda^n + B$ based on which fit has a lower reduced chi-square statistic. We compute the uncertainties (error bars) from the standard deviation of the survival probability at each sequence length. We use the uncertainties to scale the fit residuals and propagate error through each fit. The data in the $b = 0$ case have much smaller error bars due to minimal impact that transmon errors have on a single code word. The shaded regions cover the best fit with σ uncertainty in the fit parameters. The fitted offset of the exponentials has fairly high uncertainty even for the non-error detected gates. The extracted error-detected average gate fidelity is 0.9995 ± 0.0001 with non-error detected fidelity 0.989 ± 0.001. **Bottom row**: simulated cRB results with 200 averages at each sequence length using the technique described in 8.4.2 with decoherence and parameters from Table 7.1. The insets show the first 15 points in each simulation and reveal a initial fast decrease at the short gate depths due to decoherence during the readout and phase-in of the probabilistic cavity error detection.

The survival probabilities for the initial states $|0_L\rangle, |1_L\rangle$ in the $b = 0$ case have different contrasts due to our modification to decode by measuring $|1_L\rangle$. Benchmarking shots that start and end in $|1_L\rangle$ do not require a logical rotation, but shots that start in $|0_L\rangle$ require two $R_X^L(\pi)$ rotations that introduce SPAM error and reduce measurement contrast. In Fig. 3.3, we found a preparation and measurement contrast of 0.68, which is approximately the survival probability at a sequence length of one. Spectroscopy of our prepared state reveals that the populations of $|0\rangle$ and $|4\rangle$ both approach 0.5. The source of the infidelity is not clear to us, but we suspect there are errors in the relative phase between these two Fock states which can produce leakage to $|0\rangle - |4\rangle$ reducing the state preparation fidelity. Randomized benchmarking encodes gate fidelity in the decay *rate* of survival probabilities, which is separate from contrast reduction due to state preparation and measurement errors. For these reasons, we choose to fit the survival probability decay of each state separately and then average their decay rates to find λ_1, otherwise the unequal contrast propagates into unequally weighting the two decay rates.

The second potential problem is the difficulty of fitting the linear short-term behavior of an exponential. In Fig. 8.1 we saw that adding error detection changes the leakage properties of the gate, resulting in a different survival probability plateau at long gate depths. At short gate depths, we can approximate an exponential $A(\lambda^n + B) \sim A(1 + \log(\lambda)n) + AB$, but linear fits $\alpha n + \beta$ cannot distinguish the two offsets A (the exponential contrast) and AB, which impacts our estimation of $\log(\lambda) = \frac{\alpha(1+B)}{\beta}$. Furthermore, the binomial code decays through its leakage space partly back into its code space, setting different plateaus B at different timescales, as we see in Fig. 8.5. Since the gates in G_S possess similar error-detecting probabilities to those in Fig. 8.1, which plateaus with survival probability 0.5, we set $B = 0.5$ here. This value also represents a worst-case choice, as $\lambda = \exp\left(\frac{\alpha(1+B)}{\beta}\right)$ decreases with larger B since the slopes α are always negative. We make this choice when fitting the

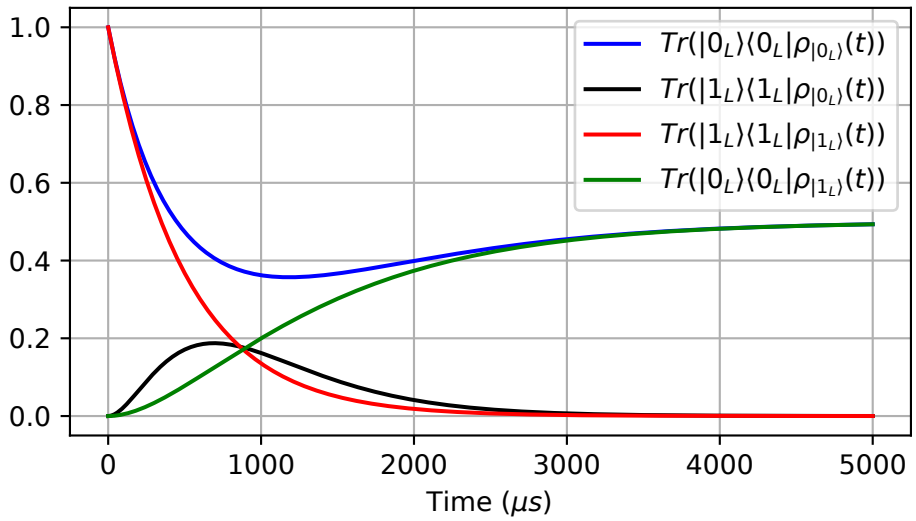


Figure 8.5: Binomial code word populations under cavity decay $\mathcal{D}[\sqrt{\kappa}C]$ with $\kappa = 1/(1ms)$. $|1_L\rangle = |2\rangle$ leaks out of the code space to $|1\rangle$ when it loses one photon. However, if it loses another photon and falls to $|0\rangle$, it leaks back in to the codespace because $|\langle 0|0_L\rangle|^2 = 1/2$. The same is true for $|0_L\rangle$, which decays through $|1_L\rangle$ towards $|0\rangle$. We see this manifest as a bump in the black curve as population leaks into, and then out of, $|1_L\rangle$. This behavior complicates randomized benchmarking, which already struggles with leakage, by adding an error channel that passes from the code space, through the leakage space, and then back into partial overlap with the code space. However, at short time scales, both decay exponentially at similar rates as there is little contribution from multiple jumps [Michael et al., 2016]. We expect randomized benchmarking to be valid in this region.

cRB in Figs. 8.4 and 8.7.

With these two solutions, we extract an average non-error detected fidelity 0.989 ± 0.001 and an error-detected fidelity to 0.9995 ± 0.0001 for gates in G_S. This error-detected fidelity exceeds the $g-f$ transmon qubit fidelity from Fig. 7.6, which should not surprise us since we are now detecting the previously dominant leakage errors. Our error detection reduces the infidelity by a factor of 22, which nearly matches the best-case prediction of 30 we made in Section 5.4 based on our measurement characterization. We believe that the probabilistic logical error detection mentioned previously plays a big part in this reduction. In the cRB simulations in Fig. 8.4, we see an initial rapid decrease in survival probability before flattening out after ~ 4 gates. This transition period occurs because the logical error detection is inefficient and only accurately detects logical errors after multiple "measurements." For single gates, logical errors that occur during the gate or error detection are likely missed, producing a lower fidelity gate. The fidelity measured here is then not the fidelity of a single gate in isolation, but the average fidelity of each gate in a sequence of 71 successful gates. Since these gates detect errors other than transmon dephasing and decay, we expect the failure rate of a sequence of gates to exceed that of a single gate. We see that this is consistently true in Fig. 8.7. Further increases in fidelity are limited by undetectable double transmon jumps and cavity decay during the final transmon readout.

8.4.4 Sweeping the transmon dephasing rate

Now that we have estimated the average gate fidelity of G_S at the natural decoherence rates of our system, we would like to see how the gate fidelity scales with the transmon dephasing and decay rates. Any scaling that we measure will be the average of all gates in G_S; we do not have a way to extract individual scalings without using IRB as in [Reinhold et al., 2020]. In this chapter, we focus on characterizing the average

gate fidelity at different dephasing rates. We don't expect the scaling to approach $O(p_{\text{phys}}^2)$ because the average scaling of ~ 1.6 will be weighed down by the identity gate's near linear scaling, as seen in Fig. 8.3. Furthermore, the error scaling in Fig. 8.3 is the most optimistic scenario where we only subject the system to one decoherence mechanism at a time. When all are present, the single-gate fidelities are limited by cavity photon loss, as seen in Fig. 8.7.

To tune the value of Γ_f^ϕ, we apply a weak readout drive to the readout resonator. The transmon is dispersively coupled to the readout resonator, so photon number shot noise in the resonator induces random frequency fluctuations of the transmon. The induced dephasing rate is

$$\Gamma_k^\phi = \frac{\kappa}{2}\Re\left(\sqrt{\left(1+\frac{i\chi_r^k}{\kappa}\right)^2 + \frac{4i\chi_r^k\bar{n}}{\Gamma}} - 1\right) \tag{8.4}$$

$$\sim \frac{\bar{n}\kappa(\chi_r^k)^2}{\kappa^2 + (\chi_r^k)^2} \tag{8.5}$$

for $\bar{n} \ll 1$ photons in a resonator with loss rate κ_r dispersively coupled to the kth transmon level with strength χ_r^k [Reinhold, 2019, Rigetti et al., 2012]. In our system, both κ and χ_r^e are approximately 1 MHz. Populating the readout resonator induces additional dephasing on all transmon states, not just $|e\rangle$ or $|f\rangle$. We only characterize the induced Γ_f^ϕ because our error detection ideally removes shots where we find the transmon in $|e\rangle$.

A weak drive on the resonator produces a small coherent state that loses photons at rate $\kappa_r\bar{n}$. In addition to inducing dephasing, this coherent state produces a deterministic dispersive shift of the transmon. To operate our gates in the presence of the additional noise produced by populating the readout, we must accurately calibrate this static frequency shift. In Fig. 8.6, we plot the induced decoherence rates and deterministic frequency shifts as a function of DAC amplitude and show that we can tune Γ_f^ϕ over a full order of magnitude.

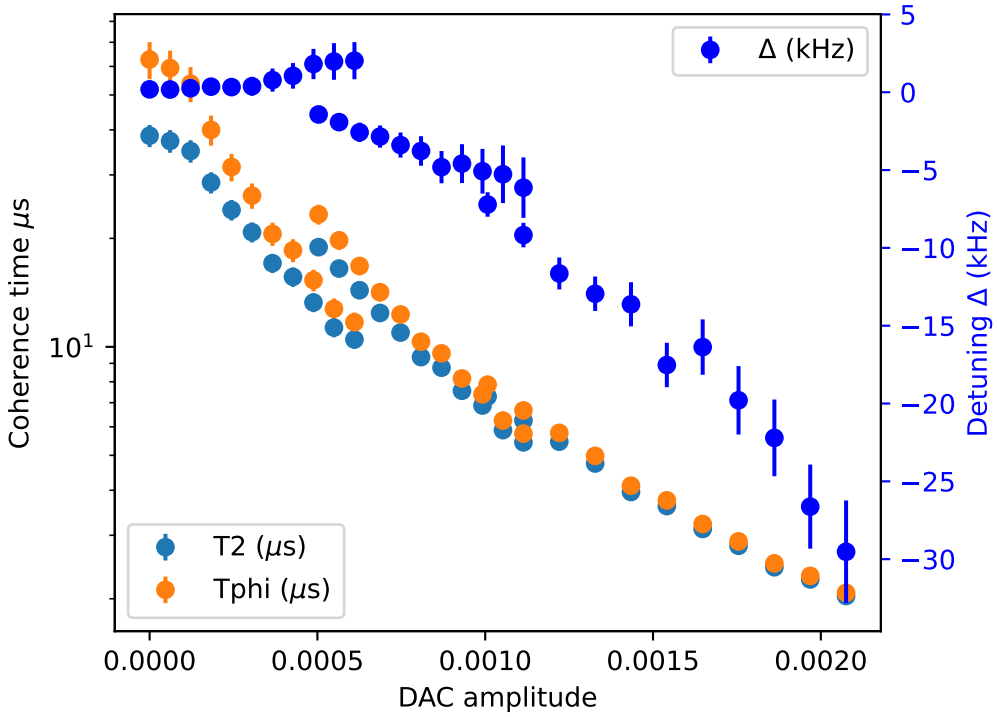


Figure 8.6: T_2^f and T_f^ϕ at various readout drive DAC amplitudes. Weak drives on the readout produce additional dephasing, but also induce a deterministic detuning (blue dots). We need to calibrate both before running cRB with artificially induced dephasing. The data here is taken in three separate groups that acquire data over different timescales. As the dephasing rate increases in each group, the fit uncertainty increases. The three groups overlap so that we can see this effect, but also acquire an accurate T_2 and Δ. The transmon used in our measurements jumped between two frequencies separated by 12.5 kHz, each with different values of T_f^ϕ at zero DAC amplitude. This data was taken when the transmon had jumped to the frequency with corresponding lower coherence times. The leftmost data points in Fig. 8.7 corresponds to the highest observed T_f^ϕ. This jumpiness means that when acquiring data for the higher coherence times, we measure T_2 and calibrate the detuning between averaging at each cRB gate depth. We check that the value of T_f^ϕ remained constant over the course of collecting survival probabilities at all gate depths after each experiment.

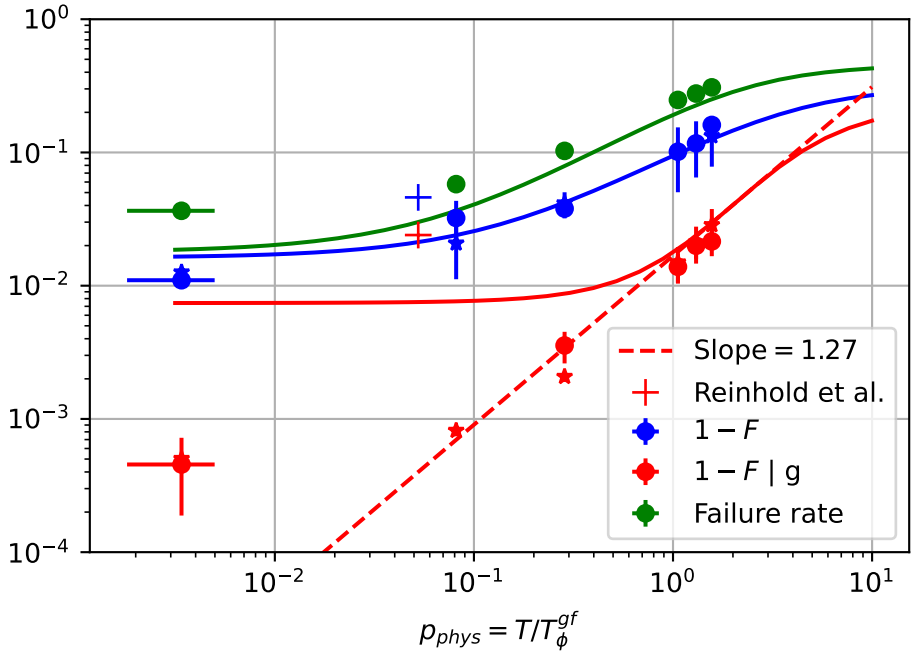


Figure 8.7: cRB results as a function of Γ_f^ϕ. Measured average gate fidelities of G_S at different values of Γ_f^ϕ set by weakly driving the readout (dots). Error bars result from the uncertainty in the T_2 measurements and the shot noise of the cRB measurements. The data with $p_{\text{phys}} \sim 0.08$ suffered from a miscalibration in the T_2 tuning sequence and did not produce a reliable error-detected fidelity. Solid lines are the average simulated fidelities of a single gate in G_S, including the error-detecting readout following each gate. The infidelity floor here is set by photon loss in the cavity, which occurs with approximate probability $p_{\text{phys}} = T\kappa_C/2 = 0.008$, where the factor of 2 accounts for the average photon number of the code words $\bar{n} = 2$. As the failure rate of each gate increases, it becomes difficult to measure the survival probability at long gate depths. The sequence depths used for each measurement are 71, 21, 21, 11, 11, 11, ordered from smallest Γ_f^ϕ to largest. The average gate fidelities are found by simulating cRB (stars) for these same sequence length agree well and are generally covered by the experimental values. We use the measurement operator $M_g = \sqrt{0.998}|g\rangle\langle g| + \sqrt{0.001}|e\rangle\langle e| + \sqrt{0.001}|f\rangle\langle f|$ with values informed by Table 5.1. The dashed line is tangent to the simulated error-detected fidelity of a single gate.

To find the average gate fidelity of G_S at any Γ_f^ϕ, we simply drive the readout resonator with the appropriate amplitude at the same time as the we play the pulses that implement each gate in G_S, with appropriate detuning. We do not include a ringdown time after each gate as we immediately measure the transmon's state for error detection, re-populating the resonator. The results at six different values of Γ_f^ϕ are shown in Fig. 8.7.

8.5 Conclusion

The results in Fig. 8.7 are the culmination of the last four chapters, and demonstrate that error detection can provide a significant improvement in fidelity. The path here required combining the relatively new ideas of error detection and path independence with incremental upgrades to previously used $g-f$ transmon control techniques [Rosenblum et al., 2018, Reinhold et al., 2020]. The gates produced are of such high fidelity that we needed to implement a new variant of randomized benchmarking, character randomized benchmarking, on bosonic qubits (for the first time, as far as we are aware).

Comparing the results with simulations, we realized that our gates contain "built-in" logical error detection, a pleasant surprise for which we did not explicitly optimize. In Fig. 6.3, we showed that the $R_Z^L(\theta)$ gates we numerically optimize share a structure with the gate construction derived in [Tsunoda et al., 2023]. This gate construction has some built-in logical error detection that was previously unused. A successful string of gates then means that no detectable transmon or logical errors have occurred, removing the need for separate, interleaved logical error measurements. This result also suggests that a better definition of gate fidelity is needed for error-detected circuits, since the error detected fidelity per gate can depend on circuit depth.

The 0.9995 error-detected fidelity that we measure with cRB compares favorably

to other error-detected gates in superconducting qubits and neutral atoms [Lu et al., 2023, Ma et al., 2023]. This infidelity is a factor of 50 smaller than a fault-tolerant version of the same gate [Reinhold et al., 2020], which has a significantly more difficult implementation that includes non-deterministic gate durations. The simplicity of the error-detected gate, which does not require real-time transmon reset, χ-matching, or preservation of logical information in the presence of an error, contributes to this improvement. Fault-tolerant implementations can likely be improved with our numerically optimized $gf/2$ control scheme, but care needs to be taken to ensure that information in the error space is preserved.

We need to note that the R_Z^L gates we demonstrated here do not form a universal qubit gate set, as opposed to the other error-detected gates mentioned above. SNAP gates combined with displacements [Krastanov et al., 2015] or squeezing enable universal control of the binomial qubit implemented here. These gates tend to be more difficult to implement, as they need to preserve or modify phases across a large spread of Fock states. It is likely possible to find waveforms that meet these requirements, but larger control bandwidths may be needed. Increasing the values of χ_f and α will ease the difficulty in finding these controls, but can only be increased to a point. However, we still believe this is a significant demonstration because it shows that we can construct gates that detect all dominant error syndromes, which will be the goal for any architecture that utilizes a transmon ancilla for any or all of its gates.

Chapter 9

Conclusion and outlook

To end this dissertation, I want to provide some perspective on our achievements and suggest additional avenues of exploration. The work presented so far has built upon the ideas and excellent work of my predecessors and I hope that some of the tools developed here find applications in future projects.

Real-time Hidden Markov Model (HMM) state estimation

We designed the multi-bit photon-number resolving measurement in Chapter 4 to sample from the output state of multi-photon quantum simulations. While highly effective, our error mitigation scheme only deconvolves errors in the resulting sampled distribution. If we want to use this measurement, or a similar multi-bit measurement, for error correction and detection, we need to optimize the single-shot fidelity. In the quantum chemistry experiment, our goal was to obtain as much information about the input state as possible, but in error detection, we need to have high confidence in the post-measurement state. We can add redundant measurements that re-sample one of the bits to improve single-shot fidelity. For example, since single photon loss is the most likely cavity error in the bitwise measurement, a second measurement of parity at the end would reveal any single photon loss during the full sequence.

Furthermore, we can perform real-time maximum likelihood state inference using our HMM. HMMs have been used to assign states in post-processing [Martinez et al., 2020], but real-time HMM state estimation will require new logic in the FPGAs that run our control systems. One potential benefit of this scheme is that the HMM state inference assigns a degree-of-belief to its estimation, allowing us to set a threshold at which we throw out or repeat a measurement in real-time. During this project, we also considered constructing an adaptive protocol where the next measurement is chosen based on the outcome of the current one. However, Sal Elder and I found that optimizing the binary decision tree underpinning this scheme is NP-complete [Hyafil and Rivest, 1976].

Optimization of path-independent gates

In Chapter 6, we used a numerical optimizer to find $R_Z(\theta)$ logical rotations on the binomial code that appear path-independent to one dephasing jump. Recovering the promise of path-independence to all orders in dephasing is likely a much more difficult task. To truly engineer a Hamiltonian of the form

$$H_{\mathrm{SNAP}} = \Omega S(\vec{\phi})|g\rangle\langle f| + \Omega S^{\dagger}(\vec{\phi})|f\rangle\langle g| \tag{9.1}$$

may require a large amount of bandwidth not available with current control systems. We have tried optimizing $R_Z(\theta)$ logical gates with Monte-Carlo methods that include all trajectories, but trajectories with more than one jump are rare and thus only make a small contribution to the gradient. We have also not found path-independent implementations of $R_X(\theta)$ and $R_Y(\theta)$ logical gates. Further theoretical exploration is ongoing [Xu et al., 2023] and I encourage numerical explorations that promote previously static interactions, such as $\chi_f|f\rangle\langle f|C^{\dagger}C$, to parametric interactions.

Further optimization of $g-f$ driving

The $g-f$ transmon has applications in two prominent proposals for erasure qubits [Kubica et al., 2022, Teoh et al., 2023]. We've demonstrated high-fidelity $g-f$ qubits and optimal control of logical bosonic qubits using the $gf/2$ interaction, but further optimization to minimize undetected errors is always needed. At the quantum hardware level, increasing the anharmonicity α immediately reduces leakage. But there are avenues for improvement in control design as well. In the optical community, STImulated Raman Adiabatic Passage (STIRAP) is a common control scheme for three level systems that aims to minimize leakage to a lossy intermediate state. Typical schemes use two optical drives near the single photon transitions, as we do in Section 7.1.1, but newer proposals [Petiziol et al., 2020] suggest adding a third direct $g-f$ coupling (the $gf/2$ interaction, in our case). We did not try this combination, but future efforts should explore counterdiabatic methods. We also encourage efforts to extend optimal control to wide bandwidths that cover each of the methods above. The main obstacle to wide-band optimization is that integrating differential equations with fast oscillating terms tends to be more difficult than the relatively narrow-band pulses we optimized here.

Built-in error detection

One of the more surprising results in this dissertation was the emergence of logical error detection in our $R_Z(\theta)$ logical gates in Chapter 8. Closer inspection of the pulse structure in Fig. 6.3 suggests that some degree of logical error detection should be present in the exponential gadget construction of [Tsunoda et al., 2023]. Built-in error detection, even if it is not single-shot, simplifies gate circuits by removing the need for some separate logical error measurements. It's not clear that built-in single-shot error detection is the optimal choice either. For example, adding single-shot error detection to the $R_Z(\theta)$ logical gates would require nearly doubling their duration, potentially

increasing the erasure rate beyond threshold due to transmon errors. Many schemes don't perform logical error detection after every gate, and probably don't need perfect error detection built in to each gate. Understanding the trade-offs and what limits the efficiency of built-in error detection is an open question.

Appendix A

Bitwise measurement appendix

A.1 Dynamic ancilla reset statistics

As discussed in the supplemental material in [Elder et al., 2020], our ancilla qubit reset protocol does not always succeed in a single attempt. This increases the duration between bits on a shot-by-shot basis. This duration is proportional to the number of reset attempts, whose probabilities shown in Fig. A.1 have a long tail, with 35.1% of resets requiring more than one attempt. The average number of reset attempts required across all Fock states prepared in Figs. 4.4 and 4.5(b) is $\bar{n}_{\text{reset}} = 2.05$. Each reset attempt has a duration of $2.244\,\mu$s so we use the average reset time $\kappa t' = 2.244\,\mu\text{s} \cdot \kappa\bar{n}_{\text{reset}} = 0.0046$ in the model.

It is tempting to remove runs via postselection that take more than a chosen threshold of reset attempts, but this introduces bias into the measurement results as the number of required reset attempts increases with the number of photons in the storage mode. Postselecting on a low threshold such as five attempts (used in [Elder et al., 2020]) removes 5% more measurements of $|15\rangle$ than $|1\rangle$. Setting the postselection threshold at 50 attempts lowers this disparity to 0.6%, but we still choose to avoid postselection to produce an unbiased measurement. A more granular

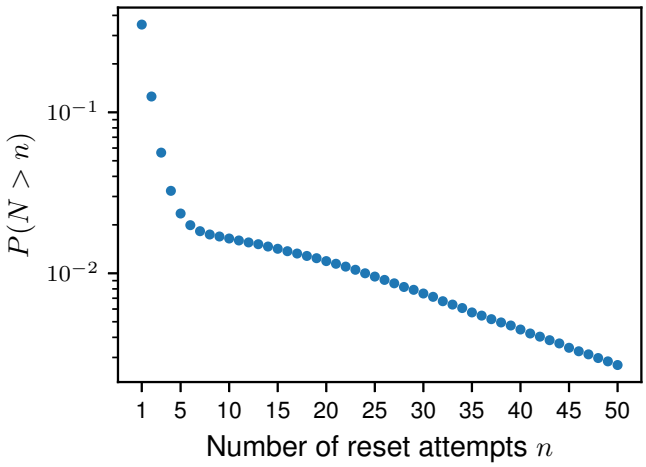


Figure A.1: Ancilla dynamic reset probabilities from the Fock state data in Figs. 4.4 and 4.5(b). After each bit measurement, we readout and, if necessary, dynamically reset the ancilla. This protocol requires more than one attempt with probability $P(N > 1) = 0.351$ and at most 50 with probability $1 - P(N > 50) = 0.997$.

approach to ancilla reset using the state-dependent reset probability distributions can be included in the HMM if one separately calibrates this dependence.

A.2 Choice of calibration states

Consider a set of calibration states $\{|\psi_j\rangle\}$. After preparing each $|\psi_j\rangle$, we measure the k^{th} bit to find the probability of measuring $b_k = 0, 1$ for each basis state $\mathbf{P}^{\text{cal}}_{b_k,j} = P(b_k = 0, 1| \, |\psi_j\rangle) = \langle\psi_j|E^{(k)}_{b_k}|\psi_j\rangle$. Our preparation protocol may suffer from transitions caused by ancilla readout and storage mode decay that we can model in the calibration process

$$\mathbf{P}^{\text{cal}}_{b_k,j} = \sum_{m,n} T_{m,n}(\kappa t^{(0)}) E^{(k)}_{b_k,n} O_{j,m}, \tag{A.1}$$

where $O_{j,m} = |\, \langle\psi_j|m\rangle \,|^2$ changes from the calibration basis $\{|\psi_j\rangle\}$ to the Fock basis. The argument of T, $\kappa t^{(0)}$, includes the duration of the final check of the preparation protocol and the $C\hat{\mathcal{P}}^0$ pulse. It can be shown that the effective transition duration for this final check is approximately half of the length of the selective pulse. The value

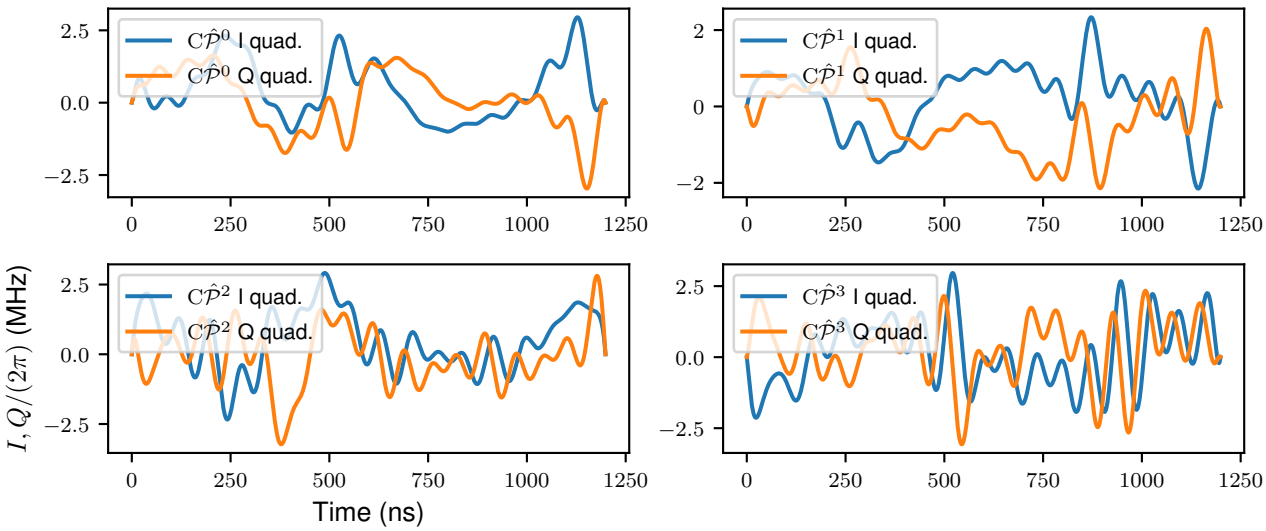


Figure A.2: $C\hat{\mathcal{P}}^k$ pulses generated by optimal control. Each pulse has a fixed duration of 1200 ns. The amplitude is multiplied by a calibrated constant before being played by FPGA-based signal synthesizers. The signal is upconverted via an IQ mixer and passed through a series of amplifiers and attenuators before reaching the ancilla qubit.

of $\kappa t^{(0)}$ is given in Table 4.1. Solving for the diagonal of $E_{b_k}^{(k)}$ we find

$$E_{b_k,i}^{(k)} = \left(T^{-1}(\kappa t^{(0)})O^{-1}\mathbf{P}^{\text{cal}}\right)_{i,i} \tag{A.2}$$

from which we average and extract $\epsilon_g^{(k)}, \epsilon_g^{(k)}$.

Our task is then to pick a basis of calibration states $|\psi_j\rangle$. In our systems, we can prepare coherent states $|\alpha\rangle$ with the highest fidelity due to the speed of preparation. If we choose a calibration basis $\{|\alpha_j\rangle\}$ such that there is very little probability finding n photons with $n > N_{\text{max}}$, we would expect these states to be an excellent candidate for a calibration basis. However, the presence of O^{-1} in Eq. A.2 complicates matters. To satisfy $n \leq N_{\text{max}}$, the range of α_j is necessarily restricted. In this regime, the coherent states are only barely linearly independent and have significant co-overlap $\langle\alpha_j|\alpha_k\rangle$. This results in O being an ill-conditioned matrix, meaning that O^{-1} magnifies small errors in $\mathbf{P}^{\text{cal}}$ [Maciejewski et al., 2020]. We avoid the ill-conditioning problem by using the Fock states $\{|j\rangle\}_{0\leq j\leq N_{\text{max}}}$ as our calibration basis resulting in $O = 1$.

Table A.1: Parameters for the Hamiltonian A.3 used in the optimal control construction of the $C\hat{\mathcal{P}}^k$ gates.

Parameter	Value
$\alpha/(2\pi)$	132 MHz
$K_C/(2\pi)$	2.59 kHz
$\chi/(2\pi)$	885 kHz
$\chi'/(2\pi)$	3.67 kHz

A.3 $C\hat{\mathcal{P}}^k$ pulse construction

We use the optimal control techniques introduced in Ref. 2.2.2 here to create the $C\hat{\mathcal{P}}^k$. The Hamiltonian for our system is

$$H = -\frac{\alpha}{2}A^\dagger A^\dagger AA - \frac{K_C}{2}C^\dagger C^\dagger CC - \chi A^\dagger AC^\dagger C + \frac{\chi'}{2}A^\dagger AC^\dagger C^\dagger CC + \epsilon_A^*(t)A + \epsilon_A(t)A^\dagger \tag{A.3}$$

with parameter values in Table A.1. We choose to optimize the pulses using the cost function in Eq. 2.57 as we do not care about scrambling the relative state phases. If we intend to use this measurement mid-circuit, we should instead use Eq. 2.55. Both quadratures of these pulses are shown in Fig. A.2 with Hamiltonian parameters listed in Table A.1. Each pulse reaches a fidelity of at least 99.9% in a lossless system.

A.4 Complete sets of commuting observables

Consider the commuting observables parity $\hat{\mathcal{P}}_0$ and super-parity $\hat{\mathcal{P}}_1$ on the subspace $N < 4$. While both observables have twice-degenerate eigenvalues $\{+1, -1\}$, the four pairings of these eigenvalues are unique and correspond to a single shared eigenstate. This can be seen by examining the spectral decomposition of the parity operator

$\hat{\mathcal{P}}_k = \hat{B}_k^0 + \hat{B}_k^1$ in terms of its measurement operators

$$\left(\hat{B}_k^{0(1)}\right)_{ii} = \begin{cases} 1(-1) & \text{if } \left\lfloor \frac{i}{2^k} \right\rfloor \text{ is even (odd)} \\ 0 & \text{else} \end{cases} \tag{A.4}$$

The spectral decomposition of successive measurements of $\hat{\mathcal{P}}_1$ and $\hat{\mathcal{P}}_0$ is

$$\hat{\mathcal{P}}_1\hat{\mathcal{P}}_0 = \hat{B}_1^0\hat{B}_0^0 + \hat{B}_1^0\hat{B}_0^1 + \hat{B}_1^1\hat{B}_0^0 + \hat{B}_1^1\hat{B}_0^1 \tag{A.5}$$

Each measurement operator $\hat{B}_1^{b_1}\hat{B}_0^{b_0}$ on the RHS is rank one, and projects into a definite photon number state (one dimensional subspace) of the $N < 4$ subspace labeled by the two measured eigenvalues. These relabeled eigenvalues b_1, b_0 are the unique binary decomposition of the measured number of photons n in the storage mode $n = \sum_k 2^{b_k}$. That is, the results of measuring $\hat{\mathcal{P}}_0$ followed by $\hat{\mathcal{P}}_1$ fully determine the post-measurement state. These sets of observables are called complete sets of commuting observables. An example of measuring a state with this CSCO is illustrated in Fig. 4.2(b).

More generally, a set of B generalized parity measurements forms a CSCO that can resolve the first 2^B photon states

$$\hat{\mathcal{P}}_{n-1}\ldots\hat{\mathcal{P}}_0 = \sum_{b_0,\ldots,b_{n-1}} \hat{B}_{n-1}^{b_n-1}\ldots\hat{B}_0^{b_0} \tag{A.6}$$

Each term on the RHS forms a measurement operator $\hat{M}_i$ that projects the system into the Fock state $|i\rangle$. Note that in principle, the measurement can be made in any order since all parity operators commute $[\hat{\mathcal{P}}_k, \hat{\mathcal{P}}_j] = 0$. This flexibility will be important when we detail the experimental implementation.

Finally, we can identify a measurement of $\hat{\mathcal{P}}_k$ as measuring the kth bit in the binary decomposition of the photon number $|n\rangle = |b_k \ldots b_0\rangle$. This technique for PNR

detection, which we refer to as “bitwise" measurement, requires the fewest binary measurements of any scheme assuming no prior knowledge of the state.

Appendix B

Transmon QNDness HMM code

Below, we insert a copy of the notebook with which we produced the state trajectories in Fig. 5.3 and transition probabilities in Table 5.1.

Transmon QND Characterization with HMM

November 6, 2023

[1]:
```
%matplotlib inline
%config InlineBackend.figure_format = 'retina'
import sys, os
from fpga_lib.dsl import Results
import h5py as h5
import numpy as np
import matplotlib.pyplot as plt
import scipy.signal
import scipy
import matplotlib.cm as cm
from sklearn import mixture
from hmmlearn import hmm
import copy
from matplotlib import colors
```

Alright, lets now import our readout trajectory data after running the circuit in Chapter 5.4 and categorize $|g\rangle$ results as 0, $|e\rangle$ as 1, and $|f\rangle$ as 2.

[2]:
```
datadir = r'/home/rsl/smeagol-data/data/exp/'
exp_name = r'control_exps.readout_exps.repeated_readouts_one'
datestamp = '20230728'
grp = '1'
fn = os.path.join(datadir, exp_name, 'archive', datestamp) + '.h5'
r = Results.create_from_file(fn, str(grp))
data = r['default'].data
sequence_length = r.run_params["repetitions"]
no_shots = r.run_params["averages_per_block"] * r.run_params["blocks_finished"]

g_se_data = r['p_g'].data
e_se_data = r['p_e'].data
f_se_data = r['p_f'].data
se_data = 0 * g_se_data + 1 * e_se_data + 2 * f_se_data

if not r.run_params['ignore_h']:
    h_se_data = r['p_h'].data
    se_data += 3 * h_se_data
```

Initialize a CategoricalHMM and make a reasonable guess at each transition, emission, and initial

probability parameters.

[3]:
```
model = hmm.CategoricalHMM(n_components=3, params='ste', init_params='',␣
 ↪n_iter=10)
model.transmat_ = np.array([[0.95, 0.05, 0],
                            [0.05, 0.94, 0.01],
                            [0, 0.10, 0.9]])
model.startprob_ = np.array([0.05,0.475,0.475])
model.emissionprob_ = np.array([[0.95, 0.05, 0],
                            [0.02, 0.95, 0.03],
                            [0.01, 0.04, 0.95]])

model.fit(se_data.reshape(-1, 1), lengths=[sequence_length] * no_shots)
model_predict = model.predict(se_data.reshape(-1, 1), lengths=[sequence_length]␣
 ↪* no_shots)
model_probs = model.predict_proba(se_data.reshape(-1, 1),␣
 ↪lengths=[sequence_length] * no_shots)
model.monitor_
```

[3]:
```
ConvergenceMonitor(
    history=[-1872518.5047480152, -1130568.0936627276, -1124463.8713005045,
-1123496.2744208237, -1123248.074949735, -1123161.7926334778,
-1123124.4626839412, -1123105.471926099, -1123094.5408239854,
-1123087.6362426819],
    iter=10,
    n_iter=10,
    tol=0.01,
    verbose=False,
)
```

[4]:
```
model.startprob_
```

[4]:
```
array([0.02402432, 0.48204098, 0.4939347 ])
```

[5]:
```
model.transmat_
```

[5]:
```
array([[0.99854562, 0.00145438, 0.        ],
       [0.08975935, 0.89276979, 0.01747086],
       [0.        , 0.13252274, 0.86747726]])
```

[6]:
```
model.emissionprob_
```

[6]:
```
array([[9.97621679e-01, 2.37832056e-03, 0.00000000e+00],
       [1.03541439e-02, 9.84310394e-01, 5.33546175e-03],
       [2.35342583e-02, 5.99134413e-04, 9.75866607e-01]])
```

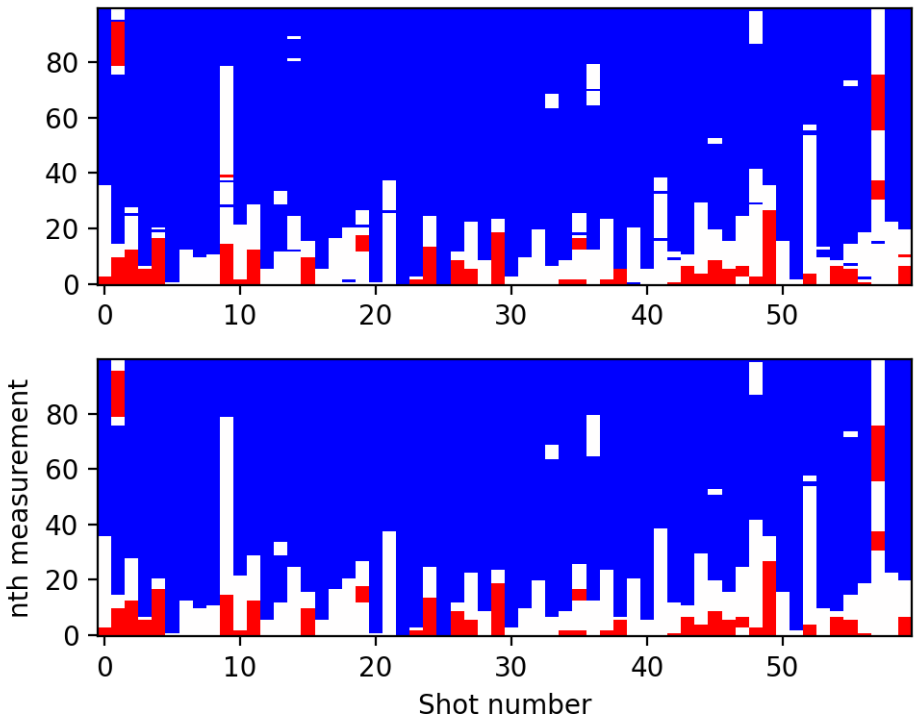
Sometimes states get mislabelled and we have to relabel them. We swap the 1 and 2 labels below

[ ]:
```
switch1 = 1
switch2 = 2

intermediate = np.where(model_predict[1] == switch1, 5, model_predict[1])
intermediate = np.where(intermediate == switch2, switch1, intermediate)
model_predict = (0, np.where(intermediate == 5, switch2, intermediate)) # keeps␣
 ↪[1] indexing below
```

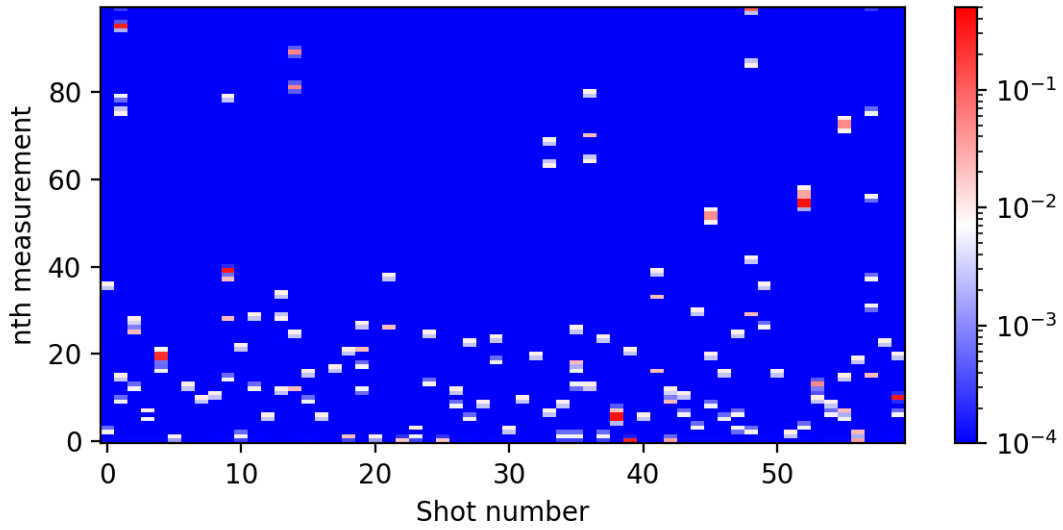
Now compare the data and trained HMM predictions. This output is the same as Fig. 5.3.

[7]:
```
fig, ax = plt.subplots(2, 1, figsize=(5, 4))
ax[0].imshow(se_data[80:140, :].T, aspect='auto', origin='lower',␣
 ↪interpolation='none', cmap='bwr')
ax[1].imshow(model_predict[8000:14000].reshape((60, 100)).T, aspect='auto',␣
 ↪origin='lower', interpolation='none', cmap='bwr')
ax[1].set_xlabel("Shot number", fontsize=10)
ax[1].set_ylabel("nth measurement", fontsize=10)
fig.tight_layout()
```



Let's now make an image of the posterior probability of each state in the prediction above. We plot 1-DOB (degree of belief) and see that the model's confidence in some state predictions is lower than others.

[8]:
```
fig, ax = plt.subplots(1, 1, figsize=(6, 3))
plot = ax.imshow(1-np.max(model_probs[8000:14000, :], axis=1).reshape((60,␣
 ↪100)).T, aspect='auto', origin='lower', interpolation='none', cmap='bwr',␣
 ↪norm=colors.LogNorm(vmin=1e-4, vmax=0.5))
ax.set_xlabel("Shot number", fontsize=10)
ax.set_ylabel("nth measurement", fontsize=10)
fig.colorbar(plot, ax=ax)
fig.tight_layout()
```



Appendix C

Additional control waveforms

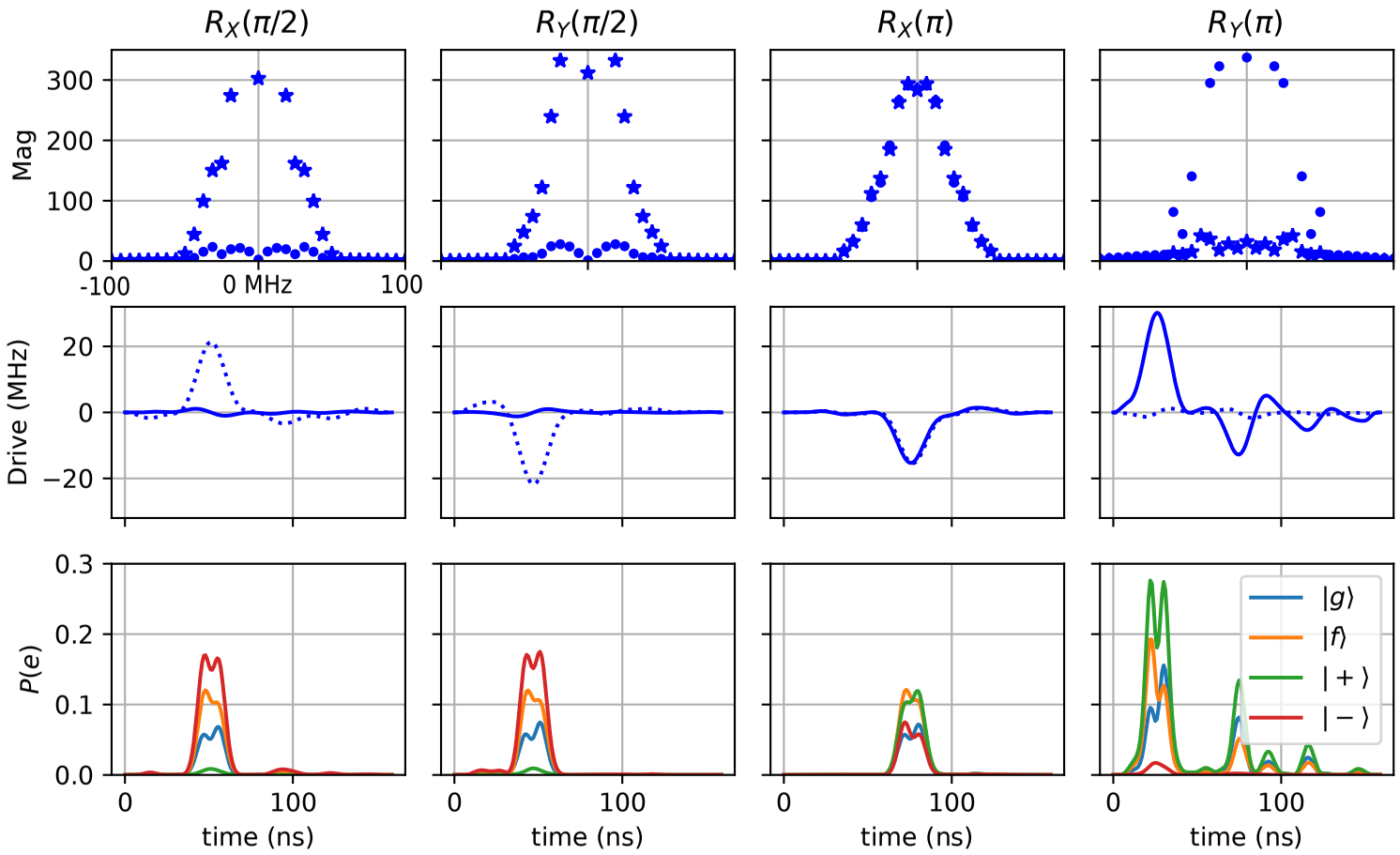


Figure C.1: Pulses are optimized with the same parameters and methods in Fig. 7.5, but with twice the length 160 ns. Intermediate leakage to $|e\rangle$ is substantially lower with these waveforms, at the cost of more incoherent leakage via excitation decay.

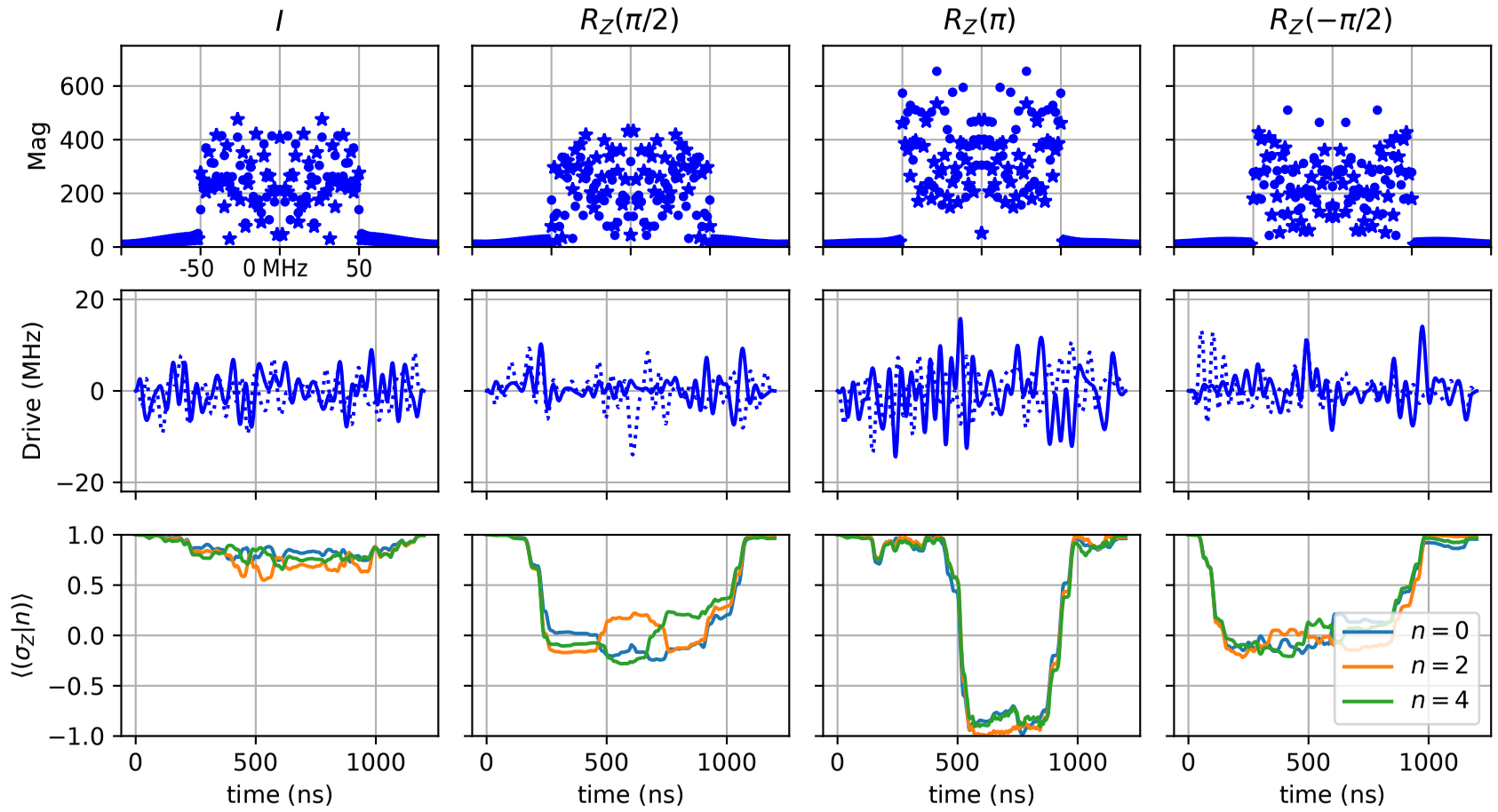


Figure C.2: Pulse waveforms used in Section 8.4. The pulses were optimized with Hamiltonian 8.3, parameters in Table 7.1, and cost function 6.53. **Top row**: Fourier amplitudes of $I(t)$ (dots) and $Q(t)$ (stars). All figures in the top row share the same x-axis. The pulses were optimized to cover 25 MHz of bandwidth, the sharp rolloff marks that boundary. **Middle row**: Time-domain waveforms of $I(t)$ (solid lines) and $Q(t)$ (dashed lines). **Bottom row**: the latitude of the transmon conditioned on each Fock state present in the binomial code.

Appendix D

Code used to generate $R_Z^L(\theta)$ gates

Below we attach the code used to generate the pulses in Figs. 8.3 and C.2. The code operates on an internal repository found at QOGS (Quantum Optimal Gate Synthesizer) written in collaboration with Alec Eickbusch. The optimizer is built on TensorFlow to enable GPU acceleration and autodifferentiation.

Optimization of Path-independent Z Rotations

November 6, 2023

[ ]:
```
%load_ext autoreload
%autoreload 2
%matplotlib inline
%config InlineBackend.figure_format = 'retina'

import tensorflow as tf
import multiprocessing
import numpy as np
import scipy as sp
import scipy.signal as signal
import qutip as qt
from qutip.qip.operations.gates import rx, ry, rz, hadamard_transform
from QOGS.optimizer.tf_adam_optimizer import AdamOptimizer
from QOGS.gate_sets import GRAPE, PI_GRAPE, trajGRAPE
import matplotlib.pyplot as plt
from matplotlib.gridspec import GridSpec
import matplotlib as mpl
import h5py
import copy
import os

opt_type = "PI" # must be "PI", "GRAPE", or "trajGRAPE"
name_prefix = 'gf_PI_'
detect_parity = False
end_in_f = False
```

[ ]:
```
### GF/2 COUPLING ###
q_dim = 5
c_dim = 5
DAC_time_resolution = 2 # in ns

# In GHz = cycles / ns
anharm = 0.134
kerr = 507e-9 # 864e-9
chi_e = 387e3 * 1e-9
chi_p_e = 400e-9
chi_f = 2 * 289.6e3 * 1e-9 # 2 is the bosonic factor
```

```
drive = D = 2 * np.pi * 1e-3

detuning = anharm / 2
e_decay_rate = 1 / 100 / 5

a = qt.tensor(qt.destroy(c_dim), qt.qeye(q_dim)) # cavity
b = qt.tensor(qt.qeye(c_dim), qt.destroy(q_dim)) # qubit
ad = a.dag()
bd = b.dag()
Igg = qt.tensor(qt.qeye(c_dim), qt.fock_dm(q_dim,0))
Iee = qt.tensor(qt.qeye(c_dim), qt.fock_dm(q_dim,1))
Iff = qt.tensor(qt.qeye(c_dim), qt.fock_dm(q_dim,2))

jump_op = qt.tensor(qt.fock_dm(c_dim, 0) + qt.fock_dm(c_dim, 2) + qt.
 ↪fock_dm(c_dim, 4), qt.fock_dm(q_dim, 2)) * np.sqrt(2 / 5000) # this is the␣
 ↪jump operator for gf dephasing in the codespace

H0 = detuning * bd * b
H0 += -(anharm/2) * bd * bd * b * b
H0 += -(kerr/2) * ad * ad * a * a
H0 += -chi_e * ad * a * qt.tensor(qt.qeye(c_dim), qt.basis(q_dim, 1).proj())
H0 += -chi_f * ad * a * qt.tensor(qt.qeye(c_dim), qt.basis(q_dim, 2).proj())
H0 += chi_p_e / 2 * ad * ad * a * a * qt.tensor(qt.qeye(c_dim), qt.basis(q_dim,␣
 ↪1).proj())
H0 *= 2*np.pi

H0 -= 1j / 2 * qt.tensor(qt.qeye(c_dim), qt.fock_dm(q_dim, 1)) * e_decay_rate
H0 -= 1j / 2 * qt.tensor(qt.qeye(c_dim), qt.fock_dm(q_dim, 3)) * e_decay_rate
H0 -= 1j / 2 * qt.tensor(qt.qeye(c_dim), qt.fock_dm(q_dim, 4)) * e_decay_rate

Hcs = [D*(b + bd), 1j*D*(b - bd)] # I, Q

success_op = qt.tensor(qt.fock_dm(c_dim, 0) + qt.fock_dm(c_dim, 2) + qt.
 ↪fock_dm(c_dim, 4), qt.fock_dm(q_dim, 2 if end_in_f else 0))
```

1 Z gate setup

[ ]:
```
# define the logical gate to apply
l0 = qt.basis(2,0) # +Z
l1 = qt.basis(2,1) # -Z
plus = (l0 + l1) / np.sqrt(2)
i_state = (l0 + 1j * l1) / np.sqrt(2)
mi_state = (l0 - 1j * l1) / np.sqrt(2)

# define the logical basis states in the basis of the system
l0_sys = (qt.basis(c_dim,0) + qt.basis(c_dim,4))/np.sqrt(2)
```

```
l1_sys = qt.basis(c_dim,2)

# define two functions that convert logical operators/states to system␣
 ↪operators/states
logical_to_sys = lambda Q: Q.full()[0,0]*l0_sys*l0_sys.dag() + Q.
 ↪full()[0,1]*l0_sys*l1_sys.dag() + Q.full()[1,0]*l1_sys*l0_sys.dag() + Q.
 ↪full()[1,1]*l1_sys*l1_sys.dag()
logical_to_sys_op = lambda Q: Q.full()[0,0]*l0_sys*l0_sys.dag() + Q.
 ↪full()[0,1]*l0_sys*l1_sys.dag() + Q.full()[1,0]*l1_sys*l0_sys.dag() + Q.
 ↪full()[1,1]*l1_sys*l1_sys.dag() + (qt.qeye(N) - l0_sys.proj() + l1_sys.
 ↪proj())
logical_to_sys_state = lambda Q: Q.full()[0, 0]*l0_sys + Q.full()[1, 0]*l1_sys␣
 ↪# Q is a vector, so the second superfluous index is always zero

# list of gate names. we'll eventually make a dictionary of gate name and␣
 ↪initial/final states
gates = ['RZ', 'RZ2', 'RZm2', 'I']
unitaries = [rz(np.pi), rz(np.pi / 2), rz(-np.pi / 2), rz(0)]
unitaries_dict = dict(zip(gates, unitaries))
to_optimize = gates

preempt_Kerr = True
preempt_angle = (-1j) * (2876 * 1 * kerr / 2 * 2 * np.pi) # note that the sign␣
 ↪is flipped because the Kerr term is negative in H. (-i) is for time evolution
if preempt_Kerr:
    pre_op = (preempt_angle * ad * ad * a * a).expm()
    name_prefix += 'Kerr_'
else:
    pre_op = qt.tensor(qt.qeye(c_dim), qt.qeye(q_dim))

# now collect gate name, initial, and final states
setups = {}
for k in to_optimize:
    init_logical_states = [l0, l1, rx(np.pi / 2) * l0, rx(-np.pi / 2) * l0]#,␣
 ↪unitaries_dict['RYm2'] * l0, unitaries_dict['RXm2'] * l0]
    final_logical_states = [unitaries_dict[k] * j for j in init_logical_states]
    init_states = [qt.tensor(logical_to_sys_state(j), qt.basis(q_dim, 0)) for j␣
 ↪in init_logical_states]
    final_states = [pre_op * qt.tensor(logical_to_sys_state(j), qt.basis(q_dim,␣
 ↪2 if end_in_f else 0)) for j in final_logical_states]
    non_PI_inits = []
    non_PI_targs = []
    name = name_prefix + 'Binomial_GRAPE_' + k
    N_blocks = 601 # try lowering to 501
    n_circuits = 0
    n_traj = 0
```

```
   n_circuits = 32
   non_PI_inits = []
   non_PI_targs = []
   PI_weight = 1.0
   synth_params = {
   'N_blocks': N_blocks, # note that the length of the pulse is this times the␣
↪DAC_time_resolution
   'N_multistart' : n_circuits, #Batch size (number of circuit optimizations␣
↪to run in parallel)
   'epochs' : 1000, #number of epochs before termination
   'epoch_size' : 5, #number of adam steps per epoch
   'learning_rate' : tf.keras.optimizers.schedules.
↪PiecewiseConstantDecay([100, 200 if opt_type=='PI' else 500], [1.0, 0.1, 0.
↪05]), #learning rate for adam 0.1 if opt_type == 'PI' else
   'term_fid' : 2.0, #0.995, #terminal fidelitiy
   'dfid_stop' : 1e-7, #stop if dfid between two epochs is smaller than this␣
↪number
   'initial_states' : init_states, #qubit tensor oscillator, start in |g> |0>
   'target_states' : final_states, #end in |e> |target>.
   'name' : name, #name for printing and saving
   'coherent' : True,
   'filename' : 'OCPs/' + name + '.h5', #if no filename specified, results␣
↪will be saved in this folder under 'name.h5'
   }

   # We initialize the gateset here
   gate_set_params = {
   'H_static' : H0,
   'H_control' : Hcs,
   'DAC_delta_t' : DAC_time_resolution,
   'inplace' : False, # true uses less memory, but is slower. Just use false
   'scale' : [400 if k == 'RZ' else 200], # range of DAC amplitudes for␣
↪initial random waves 50 for ge, 20 for gf
   'bandwidth' : [0.15], # 0.5 with leakage reduction, 0.2 otherwise
   'ringup' : 20,
   'jump_ops' : jump_op if opt_type == "PI" else [jump_op], # another annoying␣
↪API difference that I need to fix
   'non_PI_inits' : non_PI_inits,
   'non_PI_targs' : non_PI_targs,
   'success_op' : success_op,
   'threshold_start' : 0.85,
   'threshold_end' : 0.9,
   'PI_weight' : PI_weight,
   'n_trajs' : n_traj,
   'gatesynthargs': synth_params
   }
```

```
    setups[k] = gate_set_params
```

2 Optimize pulses

[ ]:
```
value = setups['RZ2']
GRAPE_gate_set = PI_GRAPE(**value)
opt = AdamOptimizer(GRAPE_gate_set)
opt.optimize()
```

Bibliography

[Abdelhafez et al., 2019] Abdelhafez, M., Schuster, D. I., and Koch, J. (2019). Gradient-based optimal control of open quantum systems using quantum trajectories and automatic differentiation. *Physical Review A*, 99(5):052327. (Cited on pages 36 and 123).

[Acharya et al., 2023] Acharya, R., Aleiner, I., Allen, R., Andersen, T. I., Ansmann, M., Arute, F., Arya, K., Asfaw, A., Atalaya, J., Babbush, R., Bacon, D., Bardin, J. C., Basso, J., Bengtsson, A., Boixo, S., Bortoli, G., Bourassa, A., Bovaird, J., Brill, L., Broughton, M., Buckley, B. B., Buell, D. A., Burger, T., Burkett, B., Bushnell, N., Chen, Y., Chen, Z., Chiaro, B., Cogan, J., Collins, R., Conner, P., Courtney, W., Crook, A. L., Curtin, B., Debroy, D. M., Del Toro Barba, A., Demura, S., Dunsworth, A., Eppens, D., Erickson, C., Faoro, L., Farhi, E., Fatemi, R., Flores Burgos, L., Forati, E., Fowler, A. G., Foxen, B., Giang, W., Gidney, C., Gilboa, D., Giustina, M., Grajales Dau, A., Gross, J. A., Habegger, S., Hamilton, M. C., Harrigan, M. P., Harrington, S. D., Higgott, O., Hilton, J., Hoffmann, M., Hong, S., Huang, T., Huff, A., Huggins, W. J., Ioffe, L. B., Isakov, S. V., Iveland, J., Jeffrey, E., Jiang, Z., Jones, C., Juhas, P., Kafri, D., Kechedzhi, K., Kelly, J., Khattar, T., Khezri, M., Kieferová, M., Kim, S., Kitaev, A., Klimov, P. V., Klots, A. R., Korotkov, A. N., Kostritsa, F., Kreikebaum, J. M., Landhuis, D., Laptev, P., Lau, K.-M., Laws, L., Lee, J., Lee, K., Lester, B. J., Lill, A., Liu, W., Locharla, A., Lucero, E., Malone, F. D., Marshall, J., Martin, O., McClean, J. R., McCourt, T., McEwen, M., Megrant, A., Meurer Costa, B., Mi, X., Miao, K. C., Mohseni, M., Montazeri, S., Morvan, A., Mount, E., Mruczkiewicz, W., Naaman, O., Neeley, M., Neill, C., Nersisyan, A., Neven, H., Newman, M., Ng, J. H., Nguyen, A., Nguyen, M., Niu, M. Y., O'Brien, T. E., Opremcak, A., Platt, J., Petukhov, A., Potter, R., Pryadko, L. P., Quintana, C., Roushan, P., Rubin, N. C., Saei, N., Sank, D., Sankaragomathi, K., Satzinger, K. J., Schurkus, H. F., Schuster, C., Shearn, M. J., Shorter, A., Shvarts, V., Skruzny, J., Smelyanskiy, V., Smith, W. C., Sterling, G., Strain, D., Szalay, M., Torres, A., Vidal, G., Villalonga, B., Vollgraff Heidweiller, C., White, T., Xing, C., Yao, Z. J., Yeh, P., Yoo, J., Young, G., Zalcman, A., Zhang, Y., Zhu, N., and Google Quantum AI (2023). Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, 614(7949):676–681. (Cited on pages 11 and 132).

[Achilles et al., 2004] Achilles, D., Silberhorn, C., Sliwa, C., Banaszek, K., Walmsley, I. A., Fitch, M. J., Jacobs, B. C., Pittman, T. B., and Franson, J. D. (2004).

Photon-number-resolving detection using time-multiplexing. *Journal of Modern Optics*, 51(9-10):1499–1515. (Cited on page 60).

[Agrawal et al., 2018] Agrawal, A., Verschueren, R., Diamond, S., and Boyd, S. (2018). A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60. (Cited on page 48).

[Andersen et al., 2020] Andersen, C. K., Remm, A., Lazar, S., Krinner, S., Lacroix, N., Norris, G. J., Gabureac, M., Eichler, C., and Wallraff, A. (2020). Repeated quantum error detection in a surface code. *Nature Physics*, 16(8):875–880. (Cited on page 132).

[Andrews et al., 2019] Andrews, R. W., Jones, C., Reed, M. D., Jones, A. M., Ha, S. D., Jura, M. P., Kerckhoff, J., Levendorf, M., Meenehan, S., Merkel, S. T., Smith, A., Sun, B., Weinstein, A. J., Rakher, M. T., Ladd, T. D., and Borselli, M. G. (2019). Quantifying error and leakage in an encoded Si/SiGe triple-dot qubit. *Nature Nanotechnology*, 14(8):747–750. (Cited on page 143).

[Ball et al., 2021] Ball, H., Biercuk, M. J., Carvalho, A. R. R., Chen, J., Hush, M., Castro, L. A. D., Li, L., Liebermann, P. J., Slatyer, H. J., Edmunds, C., Frey, V., Hempel, C., and Milne, A. (2021). Software tools for quantum control: improving quantum computer performance through noise and error suppression. *Quantum Science and Technology*, 6(4):044011. (Cited on page 32).

[Banaszek and Walmsley, 2003] Banaszek, K. and Walmsley, I. A. (2003). Photon counting with a loop detector. *Optics Letters*, 28(1):52–54. (Cited on page 60).

[Battistel et al., 2021] Battistel, F., Varbanov, B., and Terhal, B. (2021). Hardware-Efficient Leakage-Reduction Scheme for Quantum Error Correction with Superconducting Transmon Qubits. *PRX Quantum*, 2(3):030314. (Cited on page 132).

[Benioff, 1980] Benioff, P. (1980). The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *Journal of Statistical Physics*, 22(5):563–591. (Cited on page 7).

[Bennett and Brassard, 2014] Bennett, C. H. and Brassard, G. (2014). Quantum cryptography: Public key distribution and coin tossing. *Theoretical Computer Science*, 560:7–11. (Cited on page 7).

[Bentivegna et al., 2015] Bentivegna, M., Spagnolo, N., Vitelli, C., Flamini, F., Viggianiello, N., Latmiral, L., Mataloni, P., Brod, D. J., Galvão, E. F., Crespi, A., Ramponi, R., Osellame, R., and Sciarrino, F. (2015). Experimental scattershot boson sampling. *Science Advances*, 1(3):e1400255. (Cited on pages 60 and 62).

[Biamonte et al., 2017] Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., and Lloyd, S. (2017). Quantum machine learning. *Nature*, 549(7671):195–202. (Cited on page 9).

[Blais et al., 2021] Blais, A., Grimsmo, A. L., Girvin, S., and Wallraff, A. (2021). Circuit quantum electrodynamics. *Reviews of Modern Physics*, 93(2):025005. (Cited on pages 16, 17, and 19).

[Blais et al., 2004] Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. M., and Schoelkopf, R. J. (2004). Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation. *Physical Review A*, 69(6):062320. (Cited on page 16).

[Bowdrey et al., 2002] Bowdrey, M. D., Oi, D. K. L., Short, A. J., Banaszek, K., and Jones, J. A. (2002). Fidelity of Single Qubit Maps. *Physics Letters A*, 294(5-6):258–260. (Cited on pages 31, 32, 34, and 35).

[Braginsky et al., 1980] Braginsky, V. B., Vorontsov, Y. I., and Thorne, K. S. (1980). Quantum Nondemolition Measurements. *Science*, 209(4456):547–557. (Cited on page 8).

[Brune et al., 1996] Brune, M., Hagley, E., Dreyer, J., Maître, X., Maali, A., Wunderlich, C., Raimond, J. M., and Haroche, S. (1996). Observing the Progressive Decoherence of the "Meter" in a Quantum Measurement. *Physical Review Letters*, 77(24):4887–4890. (Cited on page 8).

[Brune et al., 1992] Brune, M., Haroche, S., Raimond, J. M., Davidovich, L., and Zagury, N. (1992). Manipulation of photons in a cavity by dispersive atom-field coupling: Quantum-nondemolition measurements and generation of "Schrodinger cat" states. *Physical Review A*, 45(7):5193–5214. (Cited on page 8).

[Burkhart, 2020] Burkhart, L. D. (2020). *Error-Detected Networking for 3D Circuit Quantum Electrodynamics*. PhD thesis, Yale University. (Cited on page 24).

[Cahill and Glauber, 1969] Cahill, K. E. and Glauber, R. J. (1969). Density Operators and Quasiprobability Distributions. *Physical Review*, 177(5):1882–1902. (Cited on page 47).

[Calkins et al., 2013] Calkins, B., Mennea, P. L., Lita, A. E., Metcalf, B. J., Kolthammer, W. S., Linares, A. L., Spring, J. B., Humphreys, P. C., Mirin, R. P., Gates, J. C., Smith, P. G. R., Walmsley, I. A., Gerrits, T., and Nam, S. W. (2013). High quantum-efficiency photon-number-resolving detector for photonic on-chip information processing. *Optics Express*, 21(19):22657. (Cited on page 60).

[Campagne-Ibarcq et al., 2020] Campagne-Ibarcq, P., Eickbusch, A., Touzard, S., Zalys-Geller, E., Frattini, N. E., Sivak, V. V., Reinhold, P., Puri, S., Shankar, S., Schoelkopf, R. J., Frunzio, L., Mirrahimi, M., and Devoret, M. H. (2020). Quantum error correction of a qubit encoded in grid states of an oscillator. *Nature*, 584(7821):368–372. (Cited on pages 10, 12, and 38).

[Cattaneo et al., 2018] Cattaneo, M., Paris, M. G. A., and Olivares, S. (2018). Hybrid quantum key distribution using coherent states and photon-number-resolving detectors. *Physical Review A*, 98(1):012333. (Cited on page 60).

[Chang et al., 2023] Chang, K., Singh, S., Sahah, K., Schoelkopf, R. J., Steven, G., and Puri, S. (2023). Performance of surface codes with imperfect erasure detection. *Bulletin of the American Physical Society*, Abstract: Q64.00006. (Cited on page 107).

[Chapman et al., 2023] Chapman, B. J., de Graaf, S. J., Xue, S. H., Zhang, Y., Teoh, J., Curtis, J. C., Tsunoda, T., Eickbusch, A., Read, A. P., Koottandavida, A., Mundhada, S. O., Frunzio, L., Devoret, M., Girvin, S., and Schoelkopf, R. (2023). High-On-Off-Ratio Beam-Splitter Interaction for Gates on Bosonically Encoded Qubits. *PRX Quantum*, 4(2):020355. (Cited on page 21).

[Chen et al., 2011] Chen, Y.-F., Hover, D., Sendelbach, S., Maurer, L., Merkel, S. T., Pritchett, E. J., Wilhelm, F. K., and McDermott, R. (2011). Microwave Photon Counter Based on Josephson Junctions. *Physical Review Letters*, 107(21):217401. (Cited on page 61).

[Chen et al., 2016] Chen, Z., Kelly, J., Quintana, C., Barends, R., Campbell, B., Chen, Y., Chiaro, B., Dunsworth, A., Fowler, A., Lucero, E., Jeffrey, E., Megrant, A., Mutus, J., Neeley, M., Neill, C., O'Malley, P., Roushan, P., Sank, D., Vainsencher, A., Wenner, J., White, T., Korotkov, A., and Martinis, J. M. (2016). Measuring and Suppressing Quantum State Leakage in a Superconducting Qubit. *Physical Review Letters*, 116(2):020501. (Cited on page 143).

[Chou, 2018] Chou, K. S. (2018). *Teleported operations between logical qubits in circuit quantum electrodynamics*. PhD thesis, Yale University. (Cited on pages 12, 16, and 20).

[Chou et al., 2018] Chou, K. S., Blumoff, J. Z., Wang, C. S., Reinhold, P. C., Axline, C. J., Gao, Y. Y., Frunzio, L., Devoret, M. H., Jiang, L., and Schoelkopf, R. J. (2018). Deterministic teleportation of a quantum gate between two logical qubits. *Nature*, 561(7723):368–373. (Cited on page 24).

[Chou et al., 2023] Chou, K. S., Shemma, T., McCarrick, H., Chien, T.-C., Teoh, J. D., Winkel, P., Anderson, A., Chen, J., Curtis, J., de Graaf, S. J., Garmon, J. W. O., Gudlewski, B., Kalfus, W. D., Keen, T., Khedkar, N., Lei, C. U., Liu, G., Lu, P., Lu, Y., Maiti, A., Mastalli-Kelly, L., Mehta, N., Mundhada, S. O., Narla, A., Noh, T., Tsunoda, T., Xue, S. H., Yuan, J. O., Frunzio, L., Aumentado, J., Puri, S., Girvin, S. M., Moseley, J., and Schoelkopf, R. J. (2023). Demonstrating a superconducting dual-rail cavity qubit with erasure-detected logical measurements. (Cited on page 167).

[Chow et al., 2009] Chow, J. M., Gambetta, J. M., Tornberg, L., Koch, J., Bishop, L. S., Houck, A. A., Johnson, B. R., Frunzio, L., Girvin, S. M., and Schoelkopf, R. J. (2009). Randomized Benchmarking and Process Tomography for Gate Errors in a Solid-State Qubit. *Physical Review Letters*, 102(9):090502. (Cited on page 143).

[Claes, 2023] Claes, J. (2023). (Private communication). (Cited on page 164).

[Claes et al., 2021] Claes, J., Rieffel, E., and Wang, Z. (2021). Character Randomized Benchmarking for Non-Multiplicity-Free Groups With Applications to Subspace, Leakage, and Matchgate Randomized Benchmarking. *PRX Quantum*, 2(1):010351. (Cited on pages 145 and 146).

[Clements et al., 2018] Clements, W. R., Renema, J. J., Eckstein, A., Valido, A. A., Lita, A., Gerrits, T., Nam, S. W., Kolthammer, W. S., Huh, J., and Walmsley, I. A. (2018). Approximating vibronic spectroscopy with imperfect quantum optics. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 51(24):245503. (Cited on pages 60, 63, and 85).

[Curtis et al., 2021] Curtis, J. C., Hann, C. T., Elder, S. S., Wang, C. S., Frunzio, L., Jiang, L., and Schoelkopf, R. J. (2021). Single-shot number-resolved detection of microwave photons with error mitigation. *Physical Review A*, 103(2):023705. (Cited on pages 13, 59, and 86).

[Dassonneville et al., 2020] Dassonneville, R., Assouly, R., Peronnin, T., Rouchon, P., and Huard, B. (2020). Number-Resolved Photocounter for Propagating Microwave Mode. *Physical Review Applied*, 14(4):044022. (Cited on pages 61 and 74).

[de Fouquieres et al., 2011] de Fouquieres, P., Schirmer, S. G., Glaser, S. J., and Kuprov, I. (2011). Second order gradient ascent pulse engineering. *Journal of Magnetic Resonance*, 212(2):412–417. (Cited on page 28).

[Deutsch and Penrose, 1997] Deutsch, D. and Penrose, R. (1997). Quantum theory, the Church–Turing principle and the universal quantum computer. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 400(1818):97–117. (Cited on page 7).

[Devoret et al., 1985] Devoret, M. H., Martinis, J. M., and Clarke, J. (1985). Measurements of Macroscopic Quantum Tunneling out of the Zero-Voltage State of a Current-Biased Josephson Junction. *Physical Review Letters*, 55(18):1908–1911. (Cited on page 9).

[Devoret et al., 1984] Devoret, M. H., Martinis, J. M., Esteve, D., and Clarke, J. (1984). Resonant Activation from the Zero-Voltage State of a Current-Biased Josephson Junction. *Physical Review Letters*, 53(13):1260–1263. (Cited on page 9).

[Diamond and Boyd, 2016] Diamond, S. and Boyd, S. (2016). CVXPY: A Python-Embedded Modeling Language for Convex Optimization. *Journal of machine learning research : JMLR*, 17:83. (Cited on page 48).

[Divochiy et al., 2008] Divochiy, A., Marsili, F., Bitauld, D., Gaggero, A., Leoni, R., Mattioli, F., Korneev, A., Seleznev, V., Kaurova, N., Minaeva, O., Gol'tsman, G., Lagoudakis, K. G., Benkhaoul, M., Lévy, F., and Fiore, A. (2008). Superconducting nanowire photon-number-resolving detector at telecommunication wavelengths. *Nature Photonics*, 2(5):302–306. (Cited on page 60).

[Doktorov et al., 1977] Doktorov, E. V., Malkin, I. A., and Man'ko, V. I. (1977). Dynamical symmetry of vibronic transitions in polyatomic molecules and the Franck-Condon principle. *Journal of Molecular Spectroscopy*, 64(2):302–326. (Cited on page 64).

[Dou et al., 2018] Dou, J.-P., Yang, A.-L., Du, M.-Y., Lao, D., Gao, J., Qiao, L.-F., Li, H., Pang, X.-L., Feng, Z., Tang, H., and Jin, X.-M. (2018). A broadband DLCZ quantum memory in room-temperature atoms. *Communications Physics*, 1(1):1–7. (Cited on page 59).

[Dréau et al., 2013] Dréau, A., Spinicelli, P., Maze, J. R., Roch, J.-F., and Jacques, V. (2013). Single-Shot Readout of Multiple Nuclear Spin Qubits in Diamond under Ambient Conditions. *Physical Review Letters*, 110(6):060502. (Cited on page 76).

[Duan et al., 2001] Duan, L.-M., Lukin, M. D., Cirac, J. I., and Zoller, P. (2001). Long-distance quantum communication with atomic ensembles and linear optics. *Nature*, 414(6862):413–418. (Cited on page 59).

[Duschinsky, 1937] Duschinsky, F. (1937). The Importance of the Electron Spectrum in Multi Atomic Molecules Concerning the Franck-Condon Principle. *Acta Physicochim*, URSS(7, 551). (Cited on page 64).

[Eickbusch et al., 2022] Eickbusch, A., Sivak, V., Ding, A. Z., Elder, S. S., Jha, S. R., Venkatraman, J., Royer, B., Girvin, S. M., Schoelkopf, R. J., and Devoret, M. H. (2022). Fast universal control of an oscillator with weak dispersive coupling to a qubit. *Nature Physics*, 18(12):1464–1469. (Cited on page 36).

[Eisaman et al., 2011] Eisaman, M. D., Fan, J., Migdall, A., and Polyakov, S. V. (2011). Invited Review Article: Single-photon sources and detectors. *Review of Scientific Instruments*, 82(7):071101. (Cited on pages 60 and 67).

[Ekert, 1991] Ekert, A. K. (1991). Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67(6):661–663. (Cited on page 7).

[Elder et al., 2020] Elder, S. S., Wang, C. S., Reinhold, P., Hann, C. T., Chou, K. S., Lester, B. J., Rosenblum, S., Frunzio, L., Jiang, L., and Schoelkopf, R. J. (2020). High-Fidelity Measurement of Qubits Encoded in Multilevel Superconducting Circuits. *Physical Review X*, 10(1):011001. (Cited on pages 14, 70, 71, 73, 76, 78, 79, 133, 153, and 184).

[Emerson et al., 2005] Emerson, J., Alicki, R., and Życzkowski, K. (2005). Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S347. (Cited on pages 32 and 142).

[Epstein et al., 2014] Epstein, J. M., Cross, A. W., Magesan, E., and Gambetta, J. M. (2014). Investigating the limits of randomized benchmarking protocols. *Physical Review A*, 89(6):062321. (Cited on pages 143 and 145).

[Esmaeil Zadeh et al., 2017] Esmaeil Zadeh, I., Los, J. W. N., Gourgues, R. B. M., Steinmetz, V., Bulgarini, G., Dobrovolskiy, S. M., Zwiller, V., and Dorenbos, S. N. (2017). Single-photon detectors combining high efficiency, high detection rates, and ultra-high timing resolution. *APL Photonics*, 2(11):111301. (Cited on page 60).

[Essig et al., 2021] Essig, A., Ficheux, Q., Peronnin, T., Cottet, N., Lescanne, R., Sarlette, A., Rouchon, P., Leghtas, Z., and Huard, B. (2021). Multiplexed Photon Number Measurement. *Physical Review X*, 11(3):031045. (Cited on page 61).

[Esteve et al., 1986] Esteve, D., Devoret, M. H., and Martinis, J. M. (1986). Effect of an arbitrary dissipative circuit on the quantum energy levels and tunneling of a Josephson junction. *Physical Review B*, 34(1):158–163. (Cited on page 9).

[Fales et al., 2020] Fales, B. S., Curtis, E. R., Johnson, K. G., Lahana, D., Seritan, S., Wang, Y., Weir, H., Martínez, T. J., and Hohenstein, E. G. (2020). Performance of Coupled-Cluster Singles and Doubles on Modern Stream Processing Architectures. *Journal of Chemical Theory and Computation*, 16(7):4021–4028. (Cited on page 63).

[Feynman, 1982] Feynman, R. P. (1982). Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488. (Cited on page 7).

[Fitch et al., 2003] Fitch, M. J., Jacobs, B. C., Pittman, T. B., and Franson, J. D. (2003). Photon-number resolution using time-multiplexed single-photon detectors. *Physical Review A*, 68(4):043814. (Cited on page 60).

[Fleischhauer et al., 1999] Fleischhauer, M., Unanyan, R., Shore, B. W., and Bergmann, K. (1999). Coherent population transfer beyond the adiabatic limit: Generalized matched pulses and higher-order trapping states. *Physical Review A*, 59(5):3751–3760. (Cited on page 134).

[Frattini et al., 2022] Frattini, N. E., Cortiñas, R. G., Venkatraman, J., Xiao, X., Su, Q., Lei, C. U., Chapman, B. J., Joshi, V. R., Girvin, S. M., Schoelkopf, R. J., Puri, S., and Devoret, M. H. (2022). The squeezed Kerr oscillator: spectral kissing and phase-flip robustness. (Cited on pages 10 and 38).

[Frey et al., 2017] Frey, V. M., Mavadia, S., Norris, L. M., de Ferranti, W., Lucarelli, D., Viola, L., and Biercuk, M. J. (2017). Application of optimal band-limited control protocols to quantum noise sensing | Nature Communications. *Nature Communications*, 8(1):2189. (Cited on page 31).

[Fösel et al., 2020] Fösel, T., Krastanov, S., Marquardt, F., and Jiang, L. (2020). Efficient cavity control with SNAP gates. *arXiv:2004.14256 [quant-ph]*. (Cited on pages 117 and 118).

[Gambetta et al., 2011] Gambetta, J. M., Motzoi, F., Merkel, S. T., and Wilhelm, F. K. (2011). Analytic control methods for high-fidelity unitary operations in a weakly nonlinear oscillator. *Physical Review A*, 83(1):012308. (Cited on pages 26 and 120).

[Gammelmark et al., 2013] Gammelmark, S., Julsgaard, B., and Mølmer, K. (2013). Past Quantum States of a Monitored System. *Physical Review Letters*, 111(16):160401. (Cited on page 76).

[Gammelmark et al., 2014] Gammelmark, S., Mølmer, K., Alt, W., Kampschulte, T., and Meschede, D. (2014). Hidden Markov model of atomic quantum jump dynamics in an optically probed cavity. *Physical Review A*, 89(4):043839. (Cited on page 104).

[Ganjam et al., 2023] Ganjam, S., Wang, Y., Lu, Y., Banerjee, A., Lei, C. U., Krayzman, L., Kisslinger, K., Zhou, C., Li, R., Jia, Y., Liu, M., Frunzio, L., and Schoelkopf, R. J. (2023). Surpassing millisecond coherence times in on-chip superconducting quantum memories by optimizing materials, processes, and circuit design. (Cited on page 10).

[Gao, 2018] Gao, Y. (2018). *Multi-Cavity Operations in Circuit Quantum Electrodynamics*. PhD thesis, Yale University. (Cited on pages 12 and 16).

[Gao et al., 2019] Gao, Y. Y., Lester, B. J., Chou, K. S., Frunzio, L., Devoret, M. H., Jiang, L., Girvin, S. M., and Schoelkopf, R. J. (2019). Entanglement of bosonic modes through an engineered exchange interaction. *Nature*, 566(7745):509–512. (Cited on pages 21 and 24).

[Gao et al., 2018] Gao, Y. Y., Lester, B. J., Zhang, Y., Wang, C., Rosenblum, S., Frunzio, L., Jiang, L., Girvin, S., and Schoelkopf, R. J. (2018). Programmable Interference between Two Microwave Quantum Memories. *Physical Review X*, 8(2):021073. (Cited on pages 21 and 24).

[Gasparinetti et al., 2016] Gasparinetti, S., Berger, S., Abdumalikov, A. A., Pechal, M., Filipp, S., and Wallraff, A. J. (2016). Measurement of a vacuum-induced geometric phase. *Science Advances*, 2(5):e1501732. (Cited on page 78).

[Gaubatz et al., 1988] Gaubatz, U., Rudecki, P., Becker, M., Schiemann, S., Külz, M., and Bergmann, K. (1988). Population switching between vibrational levels in molecular beams. *Chemical Physics Letters*, 149(5):463–468. (Cited on page 134).

[Gerrits et al., 2011] Gerrits, T., Thomas-Peter, N., Gates, J. C., Lita, A. E., Metcalf, B. J., Calkins, B., Tomlin, N. A., Fox, A. E., Linares, A. L., Spring, J. B., Langford, N. K., Mirin, R. P., Smith, P. G. R., Walmsley, I. A., and Nam, S. W. (2011). On-chip, photon-number-resolving, telecommunication-band detectors for scalable photonic information processing. *Physical Review A*, 84(6):060301. (Cited on page 60).

[Gertler et al., 2021] Gertler, J. M., Baker, B., Li, J., Shirol, S., Koch, J., and Wang, C. (2021). Protecting a bosonic qubit with autonomous quantum error correction | Nature. *Nature*, 590(7845):243–248. (Cited on page 10).

[Gleyzes et al., 2007] Gleyzes, S., Kuhr, S., Guerlin, C., Bernu, J., Deléglise, S., Busk Hoff, U., Brune, M., Raimond, J.-M., and Haroche, S. (2007). Quantum

jumps of light recording the birth and death of a photon in a cavity. *Nature*, 446(7133):297–300. (Cited on page 61).

[Gottesman et al., 2001] Gottesman, D., Kitaev, A., and Preskill, J. (2001). Encoding a qubit in an oscillator. *Physical Review A*, 64(1):012310. (Cited on page 10).

[Govia et al., 2014] Govia, L. C. G., Pritchett, E. J., Xu, C., Plourde, B. L. T., Vavilov, M. G., Wilhelm, F. K., and McDermott, R. (2014). High-fidelity qubit measurement with a microwave-photon counter. *Physical Review A*, 90(6):062307. (Cited on page 61).

[Grimm et al., 2020] Grimm, A., Frattini, N. E., Puri, S., Mundhada, S. O., Touzard, S., Mirrahimi, M., Girvin, S. M., Shankar, S., and Devoret, M. H. (2020). Stabilization and operation of a Kerr-cat qubit. *Nature*, 584(7820):205–209. (Cited on pages 10, 21, and 38).

[Grover, 1996] Grover, L. K. (1996). A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing*, STOC '96, pages 212–219, New York, NY, USA. Association for Computing Machinery. (Cited on page 7).

[Guerlin et al., 2007] Guerlin, C., Bernu, J., Deléglise, S., Sayrin, C., Gleyzes, S., Kuhr, S., Brune, M., Raimond, J.-M., and Haroche, S. (2007). Progressive field-state collapse and quantum non-demolition photon counting. *Nature*, 448(7156):889–893. (Cited on page 61).

[Hamilton et al., 2017] Hamilton, C. S., Kruse, R., Sansoni, L., Barkhofen, S., Silberhorn, C., and Jex, I. (2017). Gaussian Boson Sampling. *Physical Review Letters*, 119(17):170501. (Cited on pages 60 and 62).

[Hann et al., 2018] Hann, C. T., Elder, S. S., Wang, C. S., Chou, K., Schoelkopf, R. J., and Jiang, L. (2018). Robust readout of bosonic qubits in the dispersive coupling regime. *Physical Review A*, 98(2):022305. (Cited on pages 76, 77, and 104).

[Haroche and Raimond, 2006] Haroche, S. and Raimond, J.-M. (2006). *Exploring the quantum: atoms, cavities and photons*. Oxford graduate texts. Oxford University Press, Oxford ; New York. (Cited on pages 44, 45, and 67).

[Heeres et al., 2017] Heeres, R. W., Reinhold, P., Ofek, N., Frunzio, L., Jiang, L., Devoret, M. H., and Schoelkopf, R. J. (2017). Implementing a universal gate set on a logical qubit encoded in an oscillator. *Nature Communications*, 8(1):94. (Cited on pages 12, 21, 28, 31, 36, 61, and 154).

[Heeres et al., 2015] Heeres, R. W., Vlastakis, B., Holland, E., Krastanov, S., Albert, V. V., Frunzio, L., Jiang, L., and Schoelkopf, R. J. (2015). Cavity State Manipulation Using Photon-Number Selective Phase Gates. *Physical Review Letters*, 115(13):137002. (Cited on pages 61 and 118).

[Helsen et al., 2019] Helsen, J., Xue, X., Vandersypen, L. M. K., and Wehner, S. (2019). A new class of efficient randomized benchmarking protocols. *npj Quantum Information*, 5(1):1–9. (Cited on page 145).

[Hofheinz et al., 2008] Hofheinz, M., Weig, E. M., Ansmann, M., Bialczak, R. C., Lucero, E., Neeley, M., O'Connell, A. D., Wang, H., Martinis, J. M., and Cleland, A. N. (2008). Generation of Fock states in a superconducting quantum circuit. *Nature*, 454(7202):310–314. (Cited on page 61).

[Holland et al., 2015] Holland, E., Vlastakis, B., Heeres, R., Reagor, M., Vool, U., Leghtas, Z., Frunzio, L., Kirchmair, G., Devoret, M., Mirrahimi, M., and Schoelkopf, R. (2015). Single-Photon-Resolved Cross-Kerr Interaction for Autonomous Stabilization of Photon-Number States. *Physical Review Letters*, 115(18):180501. (Cited on page 10).

[Horodecki et al., 1999] Horodecki, M., Horodecki, P., and Horodecki, R. (1999). General teleportation channel, singlet fraction, and quasidistillation. *Physical Review A*, 60(3):1888–1898. (Cited on page 142).

[Houck et al., 2007] Houck, A. A., Schuster, D. I., Gambetta, J. M., Schreier, J. A., Johnson, B. R., Chow, J. M., Frunzio, L., Majer, J., Devoret, M. H., Girvin, S. M., and Schoelkopf, R. J. (2007). Generating single microwave photons in a circuit. *Nature*, 449(7160):328–331. (Cited on page 61).

[Hu et al., 2016] Hu, J.-Y., Yu, B., Jing, M.-Y., Xiao, L.-T., Jia, S.-T., Qin, G.-Q., and Long, G.-L. (2016). Experimental quantum secure direct communication with single photons. *Light: Science & Applications*, 5(9):e16144–e16144. (Cited on page 59).

[Huh et al., 2015] Huh, J., Guerreschi, G. G., Peropadre, B., McClean, J. R., and Aspuru-Guzik, A. (2015). Boson sampling for molecular vibronic spectra. *Nature Photonics*, 9(9):615–620. (Cited on pages 60 and 63).

[Huh et al., 2020] Huh, J., Kim, K., and Peropadre, B. (2020). Sampling Photons to Simulate Molecules. *Physics*, 13:97. (Cited on page 63).

[Hyafil and Rivest, 1976] Hyafil, L. and Rivest, R. L. (1976). Constructing optimal binary decision trees is NP-complete. *Information Processing Letters*, 5(1):15–17. (Cited on page 181).

[Johansson et al., 2012] Johansson, J. R., Nation, P. D., and Nori, F. (2012). QuTiP: An open-source Python framework for the dynamics of open quantum systems. *Computer Physics Communications*, 183(8):1760–1772. (Cited on page 50).

[Johansson et al., 2013] Johansson, J. R., Nation, P. D., and Nori, F. (2013). QuTiP 2: A Python framework for the dynamics of open quantum systems. *Computer Physics Communications*, 184(4):1234–1240. (Cited on page 50).

[Johnson et al., 2010] Johnson, B. R., Reed, M. D., Houck, A. A., Schuster, D. I., Bishop, L. S., Ginossar, E., Gambetta, J. M., DiCarlo, L., Frunzio, L., Girvin, S. M., and Schoelkopf, R. J. (2010). Quantum non-demolition detection of single microwave photons in a circuit. *Nature Physics*, 6(9):663–667. (Cited on pages 61 and 67).

[Jurcevic et al., 2021] Jurcevic, P., Javadi-Abhari, A., Bishop, L. S., Lauer, I., Bogorin, D. F., Brink, M., Capelluto, L., Günlük, O., Itoko, T., Kanazawa, N., Kandala, A., Keefe, G. A., Krsulich, K., Landers, W., Lewandowski, E. P., McClure, D. T., Nannicini, G., Narasgond, A., Nayfeh, H. M., Pritchett, E., Rothwell, M. B., Srinivasan, S., Sundaresan, N., Wang, C., Wei, K. X., Wood, C. J., Yau, J.-B., Zhang, E. J., Dial, O. E., Chow, J. M., and Gambetta, J. M. (2021). Demonstration of quantum volume 64 on a superconducting quantum computing system. *Quantum Science and Technology*, 6(2):025020. (Cited on page 132).

[Kamal et al., 2009] Kamal, A., Marblestone, A., and Devoret, M. (2009). Signal-to-pump back action and self-oscillation in double-pump Josephson parametric amplifier. *Physical Review B*, 79(18):184301. (Cited on page 20).

[Kandala et al., 2017] Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J. M., and Gambetta, J. M. (2017). Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246. (Cited on page 63).

[Kandala et al., 2019] Kandala, A., Temme, K., Córcoles, A. D., Mezzacapo, A., Chow, J. M., and Gambetta, J. M. (2019). Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567(7749):491–495. (Cited on pages 63, 75, and 132).

[Kardynał et al., 2008] Kardynał, B. E., Yuan, Z. L., and Shields, A. J. (2008). An avalanche-photodiode-based photon-number-resolving detector. *Nature Photonics*, 2(7):425–428. (Cited on page 60).

[Khaneja et al., 2005] Khaneja, N., Reiss, T., Kehlet, C., Schulte-Herbrüggen, T., and Glaser, S. J. (2005). Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms. *Journal of Magnetic Resonance*, 172(2):296–305. (Cited on page 28).

[Kim et al., 2023] Kim, Y., Eddins, A., Anand, S., Wei, K. X., van den Berg, E., Rosenblatt, S., Nayfeh, H., Wu, Y., Zaletel, M., Temme, K., and Kandala, A. (2023). Evidence for the utility of quantum computing before fault tolerance. *Nature*, 618(7965):500–505. (Cited on pages 10 and 132).

[Kimble, 2008] Kimble, H. J. (2008). The quantum internet. *Nature*, 453(7198):1023–1030. (Cited on page 59).

[Kingma and Ba, 2017] Kingma, D. P. and Ba, J. (2017). Adam: A Method for Stochastic Optimization. (Cited on page 36).

[Kitaev, 1997] Kitaev, A. Y. (1997). Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191. (Cited on page 10).

[Kivlichan et al., 2018] Kivlichan, I. D., McClean, J., Wiebe, N., Gidney, C., Aspuru-Guzik, A., Chan, G. K.-L., and Babbush, R. (2018). Quantum Simulation of Electronic Structure with Linear Depth and Connectivity. *Physical Review Letters*, 120(11):110501. (Cited on page 63).

[Knill et al., 2001] Knill, E., Laflamme, R., and Milburn, G. J. (2001). A scheme for efficient quantum computation with linear optics. *Nature*, 409(6816):46–52. (Cited on pages 59 and 60).

[Knill et al., 2008] Knill, E., Leibfried, D., Reichle, R., Britton, J., Blakestad, R. B., Jost, J. D., Langer, C., Ozeri, R., Seidelin, S., and Wineland, D. J. (2008). Randomized benchmarking of quantum gates. *Physical Review A*, 77(1):012307. (Cited on page 142).

[Koch et al., 2007] Koch, J., Yu, T. M., Gambetta, J., Houck, A. A., Schuster, D. I., Majer, J., Blais, A., Devoret, M. H., Girvin, S. M., and Schoelkopf, R. J. (2007). Charge-insensitive qubit design derived from the Cooper pair box. *Physical Review A*, 76(4):042319. (Cited on pages 9, 16, 20, and 109).

[Kok et al., 2007] Kok, P., Munro, W. J., Nemoto, K., Ralph, T. C., Dowling, J. P., and Milburn, G. J. (2007). Linear optical quantum computing with photonic qubits. *Reviews of Modern Physics*, 79(1):135–174. (Cited on page 59).

[Krastanov et al., 2015] Krastanov, S., Albert, V. V., Shen, C., Zou, C.-L., Heeres, R. W., Vlastakis, B., Schoelkopf, R. J., and Jiang, L. (2015). Universal control of an oscillator with dispersive coupling to a qubit. *Physical Review A*, 92(4):040303. (Cited on pages 27, 114, 117, and 179).

[Kruse et al., 2019] Kruse, R., Hamilton, C. S., Sansoni, L., Barkhofen, S., Silberhorn, C., and Jex, I. (2019). Detailed study of Gaussian boson sampling. *Physical Review A*, 100(3):032326. (Cited on pages 60 and 62).

[Kubica et al., 2022] Kubica, A., Haim, A., Vaknin, Y., Brandão, F., and Retzker, A. (2022). Erasure qubits: Overcoming the T_1 limit in superconducting circuits. (Cited on pages 13, 90, 133, 150, and 182).

[Kurpiers et al., 2019] Kurpiers, P., Pechal, M., Royer, B., Magnard, P., Walter, T., Heinsoo, J., Salathé, Y., Akin, A., Storz, S., Besse, J.-C., Gasparinetti, S., Blais, A., and Wallraff, A. (2019). Quantum Communication with Time-Bin Encoded Microwave Photons. *Physical Review Applied*, 12(4):044067. (Cited on page 24).

[Lacroix et al., 2023] Lacroix, N., Hofele, L., Remm, A., Benhayoune-Khadraoui, O., McDonald, A., Shillito, R., Lazar, S., Hellings, C., Swiadek, F., Colao-Zanuz, D., Flasby, A., Panah, M. B., Kerschbaum, M., Norris, G. J., Blais, A., Wallraff, A., and Krinner, S. (2023). Fast Flux-Activated Leakage Reduction for Superconducting Quantum Circuits. (Cited on page 132).

[Landgraf et al., 2023] Landgraf, J., Flühmann, C., Fösel, T., Marquardt, F., and Schoelkopf, R. J. (2023). Fast quantum control of cavities using an improved protocol without coherent errors. (Cited on page 120).

[Leghtas et al., 2013] Leghtas, Z., Kirchmair, G., Vlastakis, B., Schoelkopf, R. J., Devoret, M. H., and Mirrahimi, M. (2013). Hardware-Efficient Autonomous Quantum Memory Protection. *Physical Review Letters*, 111(12):120501. (Cited on page 10).

[Leghtas et al., 2015] Leghtas, Z., Touzard, S., Pop, I. M., Kou, A., Vlastakis, B., Petrenko, A., Sliwa, K. M., Narla, A., Shankar, S., Hatridge, M. J., Reagor, M., Frunzio, L., Schoelkopf, R. J., Mirrahimi, M., and Devoret, M. H. (2015). Confining the state of light to a quantum manifold by engineered two-photon loss. *Science*, 347(6224):853–857. (Cited on page 10).

[Lett et al., 1988] Lett, P. D., Watts, R. N., Westbrook, C. I., Phillips, W. D., Gould, P. L., and Metcalf, H. J. (1988). Observation of Atoms Laser Cooled below the Doppler Limit. *Physical Review Letters*, 61(2):169–172. (Cited on page 8).

[Leung et al., 2017] Leung, N., Abdelhafez, M., Koch, J., and Schuster, D. (2017). Speedup for quantum optimal control from automatic differentiation based on graphics processing units. *Physical Review A*, 95(4):042318. (Cited on page 36).

[Li and Benjamin, 2017] Li, Y. and Benjamin, S. C. (2017). Efficient Variational Quantum Simulator Incorporating Active Error Minimization. *Physical Review X*, 7(2):021050. (Cited on page 75).

[Liao et al., 2017] Liao, S.-K., Cai, W.-Q., Liu, W.-Y., Zhang, L., Li, Y., Ren, J.-G., Yin, J., Shen, Q., Cao, Y., Li, Z.-P., Li, F.-Z., Chen, X.-W., Sun, L.-H., Jia, J.-J., Wu, J.-C., Jiang, X.-J., Wang, J.-F., Huang, Y.-M., Wang, Q., Zhou, Y.-L., Deng, L., Xi, T., Ma, L., Hu, T., Zhang, Q., Chen, Y.-A., Liu, N.-L., Wang, X.-B., Zhu, Z.-C., Lu, C.-Y., Shu, R., Peng, C.-Z., Wang, J.-Y., and Pan, J.-W. (2017). Satellite-to-ground quantum key distribution. *Nature*, 549(7670):43–47. (Cited on page 59).

[Liu and Nocedal, 1989] Liu, D. C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1):503–528. (Cited on page 36).

[Lo et al., 2012] Lo, H.-K., Curty, M., and Qi, B. (2012). Measurement-Device-Independent Quantum Key Distribution. *Physical Review Letters*, 108(13):130503. (Cited on page 59).

[Lu et al., 2023] Lu, Y., Maiti, A., Garmon, J. W. O., Ganjam, S., Zhang, Y., Claes, J., Frunzio, L., Girvin, S. M., and Schoelkopf, R. J. (2023). High-fidelity parametric beamsplitting with a parity-protected converter. *Nature Communications*, 14(1):5767. (Cited on pages 21 and 179).

[Lund et al., 2014] Lund, A., Laing, A., Rahimi-Keshari, S., Rudolph, T., O'Brien, J., and Ralph, T. (2014). Boson Sampling from a Gaussian State. *Physical Review Letters*, 113(10):100502. (Cited on page 62).

[Lundeen et al., 2009] Lundeen, J. S., Feito, A., Coldenstrodt-Ronge, H., Pregnell, K. L., Silberhorn, C., Ralph, T. C., Eisert, J., Plenio, M. B., and Walmsley, I. A. (2009). Tomography of quantum detectors. *Nature Physics*, 5(1):27–30. (Cited on page 84).

[Ma et al., 2023] Ma, S., Liu, G., Peng, P., Zhang, B., Jandura, S., Claes, J., Burgers, A. P., Pupillo, G., Puri, S., and Thompson, J. D. (2023). High-fidelity gates and mid-circuit erasure conversion in an atomic qubit. *Nature*, 622(7982):279–284. (Cited on page 179).

[Ma et al., 2022] Ma, W.-L., Li, S.-S., and Jiang, L. (2022). Algebraic structure of path-independent quantum control. *Physical Review Research*, 4(2):023102. (Cited on page 114).

[Ma et al., 2020] Ma, W.-L., Zhang, M., Wong, Y., Noh, K., Rosenblum, S., Reinhold, P., Schoelkopf, R. J., and Jiang, L. (2020). Path-Independent Quantum Gates with Noisy Ancilla. *Physical Review Letters*, 125(11):110503. (Cited on pages 91, 93, 112, 113, and 114).

[Maciejewski et al., 2020] Maciejewski, F. B., Zimborás, Z., and Oszmaniec, M. (2020). Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. *Quantum*, 4:257. (Cited on pages 75, 81, 83, and 186).

[Magesan et al., 2011] Magesan, E., Gambetta, J. M., and Emerson, J. (2011). Scalable and Robust Randomized Benchmarking of Quantum Processes. *Physical Review Letters*, 106(18):180504. (Cited on pages 142 and 143).

[Magesan et al., 2012] Magesan, E., Gambetta, J. M., Johnson, B. R., Ryan, C. A., Chow, J. M., Merkel, S. T., da Silva, M. P., Keefe, G. A., Rothwell, M. B., Ohki, T. A., Ketchen, M. B., and Steffen, M. (2012). Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking. *Physical Review Letters*, 109(8):080505. (Cited on page 145).

[Marsili et al., 2013] Marsili, F., Verma, V. B., Stern, J. A., Harrington, S., Lita, A. E., Gerrits, T., Vayshenker, I., Baek, B., Shaw, M. D., Mirin, R. P., and Nam, S. W. (2013). Detecting single infrared photons with 93% system efficiency. *Nature Photonics*, 7(3):210–214. (Cited on page 60).

[Martinez et al., 2020] Martinez, L. A., Rosen, Y. J., and DuBois, J. L. (2020). Improving qubit readout with hidden Markov models. *Physical Review A*, 102(6):062426. (Cited on pages 76, 104, and 181).

[Martinis et al., 1985] Martinis, J. M., Devoret, M. H., and Clarke, J. (1985). Energy-Level Quantization in the Zero-Voltage State of a Current-Biased Josephson Junction. *Physical Review Letters*, 55(15):1543–1546. (Cited on page 9).

[Martinis et al., 1987] Martinis, J. M., Devoret, M. H., and Clarke, J. (1987). Experimental tests for the quantum behavior of a macroscopic degree of freedom: The phase difference across a Josephson junction. *Physical Review B*, 35(10):4682–4698. (Cited on page 9).

[Mattioli et al., 2016] Mattioli, F., Zhou, Z., Gaggero, A., Gaudio, R., Leoni, R., and Fiore, A. (2016). Photon-counting and analog operation of a 24-pixel photon number resolving detector based on superconducting nanowires. *Optics Express*, 24(8):9067–9076. (Cited on page 60).

[McEwen et al., 2021] McEwen, M., Kafri, D., Chen, Z., Atalaya, J., Satzinger, K. J., Quintana, C., Klimov, P. V., Sank, D., Gidney, C., Fowler, A. G., Arute, F., Arya, K., Buckley, B., Burkett, B., Bushnell, N., Chiaro, B., Collins, R., Demura, S., Dunsworth, A., Erickson, C., Foxen, B., Giustina, M., Huang, T., Hong, S., Jeffrey, E., Kim, S., Kechedzhi, K., Kostritsa, F., Laptev, P., Megrant, A., Mi, X., Mutus, J., Naaman, O., Neeley, M., Neill, C., Niu, M., Paler, A., Redd, N., Roushan, P., White, T. C., Yao, J., Yeh, P., Zalcman, A., Chen, Y., Smelyanskiy, V. N., Martinis, J. M., Neven, H., Kelly, J., Korotkov, A. N., Petukhov, A. G., and Barends, R. (2021). Removing leakage-induced correlated errors in superconducting quantum error correction. *Nature Communications*, 12(1):1761. (Cited on page 132).

[McKay et al., 2017] McKay, D. C., Wood, C. J., Sheldon, S., Chow, J. M., and Gambetta, J. M. (2017). Efficient Z gates for quantum computing. *Physical Review A*, 96(2):022330. (Cited on page 143).

[Michael et al., 2016] Michael, M. H., Silveri, M., Brierley, R., Albert, V. V., Salmilehto, J., Jiang, L., and Girvin, S. (2016). New Class of Quantum Error-Correcting Codes for a Bosonic Mode. *Physical Review X*, 6(3):031006. (Cited on pages 90, 92, 93, 116, 162, and 173).

[Migdall et al., 1985] Migdall, A. L., Prodan, J. V., Phillips, W. D., Bergeman, T. H., and Metcalf, H. J. (1985). First Observation of Magnetically Trapped Neutral Atoms. *Physical Review Letters*, 54(24):2596–2599. (Cited on page 8).

[Milul et al., 2023] Milul, O., Guttel, B., Goldblatt, U., Hazanov, S., Joshi, L. M., Chausovsky, D., Kahn, N., Çiftyürek, E., Lafont, F., and Rosenblum, S. (2023). Superconducting Cavity Qubit with Tens of Milliseconds Single-Photon Coherence Time. *PRX Quantum*, 4(3):030336. (Cited on page 39).

[Minev et al., 2021] Minev, Z. K., Leghtas, Z., Mundhada, S. O., Christakis, L., Pop, I. M., and Devoret, M. H. (2021). Energy-participation quantization of Josephson circuits. *npj Quantum Information*, 7(1):1–11. (Cited on page 18).

[Mirrahimi et al., 2014] Mirrahimi, M., Leghtas, Z., Albert, V. V., Touzard, S., Schoelkopf, R. J., Jiang, L., and Devoret, M. H. (2014). Dynamically protected cat-qubits: a new paradigm for universal quantum computation. *New Journal of Physics*, 16(4):045014. (Cited on page 93).

[Monroe et al., 1995] Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M., and Wineland, D. J. (1995). Demonstration of a Fundamental Quantum Logic Gate. *Physical Review Letters*, 75(25):4714–4717. (Cited on page 8).

[Motzoi et al., 2009] Motzoi, F., Gambetta, J. M., Rebentrost, P., and Wilhelm, F. K. (2009). Simple Pulses for Elimination of Leakage in Weakly Nonlinear Qubits. *Physical Review Letters*, 103(11):110501. (Cited on pages 25 and 26).

[Motzoi and Wilhelm, 2013] Motzoi, F. and Wilhelm, F. K. (2013). Improving frequency selection of driven pulses using derivative-based transition suppression. *Physical Review A*, 88(6):062318. (Cited on page 26).

[Nakamura et al., 1999] Nakamura, Y., Pashkin, Y. A., and Tsai, J. S. (1999). Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature*, 398(6730):786–788. (Cited on page 9).

[Narla et al., 2016] Narla, A., Shankar, S., Hatridge, M., Leghtas, Z., Sliwa, K., Zalys-Geller, E., Mundhada, S., Pfaff, W., Frunzio, L., Schoelkopf, R., and Devoret, M. (2016). Robust Concurrent Remote Entanglement Between Two Superconducting Qubits. *Physical Review X*, 6(3):031036. (Cited on page 24).

[Ng and Tsang, 2014] Ng, S. and Tsang, M. (2014). Optimal signal processing for continuous qubit readout. *Physical Review A*, 90(2):022325. (Cited on page 76).

[Ni et al., 2023] Ni, Z., Li, S., Deng, X., Cai, Y., Zhang, L., Wang, W., Yang, Z.-B., Yu, H., Yan, F., Liu, S., Zou, C.-L., Sun, L., Zheng, S.-B., Xu, Y., and Yu, D. (2023). Beating the break-even point with a discrete-variable-encoded logical qubit. *Nature*, 616(7955):56–60. (Cited on page 10).

[Nielsen, 2002] Nielsen, M. A. (2002). A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303(4):249–252. (Cited on pages 31, 32, 34, and 35).

[Nielsen and Chuang, 2011] Nielsen, M. A. and Chuang, I. L. (2011). *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, USA, 10th edition. (Cited on pages 44 and 75).

[Nigg et al., 2012] Nigg, S. E., Paik, H., Vlastakis, B., Kirchmair, G., Shankar, S., Frunzio, L., Devoret, M. H., Schoelkopf, R. J., and Girvin, S. M. (2012). Black-Box Superconducting Circuit Quantization. *Physical Review Letters*, 108(24):240502. (Cited on page 18).

[Norris et al., 2018] Norris, L. M., Lucarelli, D., Frey, V. M., Mavadia, S., Biercuk, M. J., and Viola, L. (2018). Optimally band-limited spectroscopy of control noise using a qubit sensor. *Physical Review A*, 98(3):032315. (Cited on page 31).

[Northup and Blatt, 2014] Northup, T. E. and Blatt, R. (2014). Quantum information transfer using photons. *Nature Photonics*, 8(5):356–363. (Cited on page 59).

[Ofek et al., 2016] Ofek, N., Petrenko, A., Heeres, R., Reinhold, P., Leghtas, Z., Vlastakis, B., Liu, Y., Frunzio, L., Girvin, S. M., Jiang, L., Mirrahimi, M., Devoret, M. H., and Schoelkopf, R. J. (2016). Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature*, 536(7617):441–445. (Cited on page 10).

[Opremcak et al., 2018] Opremcak, A., Pechenezhskiy, I. V., Howington, C., Christensen, B. G., Beck, M. A., Leonard, E., Suttle, J., Wilen, C., Nesterov, K. N., Ribeill, G. J., Thorbeck, T., Schlenker, F., Vavilov, M. G., Plourde, B. L. T., and McDermott, R. (2018). Measurement of a superconducting qubit with a microwave photon counter. *Science*, 361(6408):1239–1242. (Cited on page 61).

[Paul, 1990] Paul, W. (1990). Electromagnetic traps for charged and neutral particles. *Reviews of Modern Physics*, 62(3):531–540. (Cited on page 7).

[Paul and Steinwedel, 1953] Paul, W. and Steinwedel, H. (1953). Notizen: Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforschung A*, 8(7):448–450. (Cited on page 7).

[Petiziol et al., 2020] Petiziol, F., Arimondo, E., Giannelli, L., Mintert, F., and Wimberger, S. (2020). Optimized three-level quantum transfers based on frequency-modulated optical excitations | Scientific Reports. *Scientific Reports*, 10(1):2185. (Cited on page 182).

[Pfaff et al., 2017] Pfaff, W., Axline, C. J., Burkhart, L. D., Vool, U., Reinhold, P., Frunzio, L., Jiang, L., Devoret, M. H., and Schoelkopf, R. J. (2017). Controlled release of multiphoton quantum states from a microwave cavity memory. *Nature Physics*, 13(9):882–887. (Cited on page 24).

[Puri et al., 2017] Puri, S., Boutin, S., and Blais, A. (2017). Engineering the quantum states of light in a Kerr-nonlinear resonator by two-photon driving. *npj Quantum Information*, 3(1):1–7. (Cited on pages 10 and 38).

[Puri et al., 2020] Puri, S., St-Jean, L., Gross, J. A., Grimm, A., Frattini, N. E., Iyer, P. S., Krishna, A., Touzard, S., Jiang, L., Blais, A., Flammia, S. T., and Girvin, S. M. (2020). Bias-preserving gates with stabilized cat qubits. *Science Advances*, 6(34):eaay5901. (Cited on page 10).

[Reagor et al., 2016] Reagor, M., Pfaff, W., Axline, C., Heeres, R. W., Ofek, N., Sliwa, K., Holland, E., Wang, C., Blumoff, J., Chou, K., Hatridge, M. J., Frunzio, L., Devoret, M. H., Jiang, L., and Schoelkopf, R. J. (2016). Quantum memory with

millisecond coherence in circuit QED. *Physical Review B*, 94(1):014506. (Cited on pages 10 and 39).

[Reinhold, 2019] Reinhold, P. (2019). *Controlling Error-Correctable Bosonic Qubits*. PhD thesis, Yale University. (Cited on pages 12, 21, 36, 148, and 175).

[Reinhold et al., 2020] Reinhold, P., Rosenblum, S., Ma, W.-L., Frunzio, L., Jiang, L., and Schoelkopf, R. J. (2020). Error-corrected gates on an encoded qubit. *Nature Physics*, 16(8):822–826. (Cited on pages 13, 14, 21, 27, 70, 110, 114, 118, 133, 135, 154, 159, 162, 174, 178, and 179).

[Rigetti et al., 2012] Rigetti, C., Gambetta, J. M., Poletto, S., Plourde, B. L. T., Chow, J. M., Córcoles, A. D., Smolin, J. A., Merkel, S. T., Rozen, J. R., Keefe, G. A., Rothwell, M. B., Ketchen, M. B., and Steffen, M. (2012). Superconducting qubit in a waveguide cavity with a coherence time approaching 0.1 ms. *Physical Review B*, 86(10):100506. (Cited on page 175).

[Rol et al., 2017] Rol, M., Bultink, C., O'Brien, T., de Jong, S., Theis, L., Fu, X., Luthi, F., Vermeulen, R., de Sterke, J., Bruno, A., Deurloo, D., Schouten, R., Wilhelm, F., and DiCarlo, L. (2017). Restless Tuneup of High-Fidelity Qubit Gates. *Physical Review Applied*, 7(4):041001. (Cited on page 148).

[Rosenblum et al., 2018] Rosenblum, S., Reinhold, P., Mirrahimi, M., Jiang, L., Frunzio, L., and Schoelkopf, R. J. (2018). Fault-tolerant detection of a quantum error. *Science*, 361(6399):266–270. (Cited on pages 13, 14, 21, 39, 133, 135, and 178).

[Ryan et al., 2009] Ryan, C. A., Laforest, M., and Laflamme, R. (2009). Randomized benchmarking of single- and multi-qubit control in liquid-state NMR quantum information processing. *New Journal of Physics*, 11(1):013034. (Cited on page 143).

[Sank et al., 2016] Sank, D., Chen, Z., Khezri, M., Kelly, J., Barends, R., Campbell, B., Chen, Y., Chiaro, B., Dunsworth, A., Fowler, A., Jeffrey, E., Lucero, E., Megrant, A., Mutus, J., Neeley, M., Neill, C., O'Malley, P., Quintana, C., Roushan, P., Vainsencher, A., White, T., Wenner, J., Korotkov, A. N., and Martinis, J. M. (2016). Measurement-Induced State Transitions in a Superconducting Qubit: Beyond the Rotating Wave Approximation. *Physical Review Letters*, 117(19):190503. (Cited on page 20).

[Schilpp, 1991] Schilpp, P. A. (1991). *Albert Einstein: philosopher-scientist*. Number 7 in The library of living philosophers. Open court, La Salle (Ill.). (Cited on page 7).

[Schulte-Herbrüggen et al., 2011] Schulte-Herbrüggen, T., Spörl, A., Khaneja, N., and Glaser, S. J. (2011). Optimal control for generating quantum gates in open dissipative systems. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 44(15):154013. (Cited on pages 32 and 33).

[Schuster, 2007] Schuster, D. I. (2007). *Circuit Quantum Electrodynamics*. PhD thesis, Yale University. (Cited on pages 12 and 16).

[Schuster et al., 2007] Schuster, D. I., Houck, A. A., Schreier, J. A., Wallraff, A., Gambetta, J. M., Blais, A., Frunzio, L., Majer, J., Johnson, B., Devoret, M. H., Girvin, S. M., and Schoelkopf, R. J. (2007). Resolving photon number states in a superconducting circuit. *Nature*, 445(7127):515–518. (Cited on pages 27 and 61).

[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423. (Cited on page 80).

[Shor, 1994] Shor, P. (1994). Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134. (Cited on pages 7 and 9).

[Shor, 1995] Shor, P. W. (1995). Scheme for reducing decoherence in quantum computer memory. *Physical Review A*, 52(4):R2493–R2496. (Cited on page 10).

[Simon et al., 2007] Simon, C., de Riedmatten, H., Afzelius, M., Sangouard, N., Zbinden, H., and Gisin, N. (2007). Quantum Repeaters with Photon Pair Sources and Multimode Memories. *Physical Review Letters*, 98(19):190503. (Cited on page 60).

[Sivak et al., 2023] Sivak, V. V., Eickbusch, A., Royer, B., Singh, S., Tsioutsios, I., Ganjam, S., Miano, A., Brock, B. L., Ding, A. Z., Frunzio, L., Girvin, S. M., Schoelkopf, R. J., and Devoret, M. H. (2023). Real-time quantum error correction beyond break-even | Nature. *Nature*, 616(7955):50–55. (Cited on pages 10 and 38).

[Song et al., 2022] Song, Y., Li, J., Hai, Y.-J., Guo, Q., and Deng, X.-H. (2022). Optimizing quantum control pulses with complex constraints and few variables through autodifferentiation. *Physical Review A*, 105(1):012616. (Cited on page 36).

[Sparrow et al., 2018] Sparrow, C., Martín-López, E., Maraviglia, N., Neville, A., Harrold, C., Carolan, J., Joglekar, Y. N., Hashimoto, T., Matsuda, N., O'Brien, J. L., Tew, D. P., and Laing, A. (2018). Simulating the vibrational quantum dynamics of molecules using photonics. *Nature*, 557(7707):660–667. (Cited on page 60).

[Spring et al., 2013] Spring, J. B., Metcalf, B. J., Humphreys, P. C., Kolthammer, W. S., Jin, X.-M., Barbieri, M., Datta, A., Thomas-Peter, N., Langford, N. K., Kundys, D., Gates, J. C., Smith, B. J., Smith, P. G. R., and Walmsley, I. A. (2013). Boson Sampling on a Photonic Chip. *Science*, 339(6121):798–801. (Cited on page 60).

[Sun et al., 2014] Sun, L., Petrenko, A., Leghtas, Z., Vlastakis, B., Kirchmair, G., Sliwa, K. M., Narla, A., Hatridge, M., Shankar, S., Blumoff, J., Frunzio, L., Mirrahimi, M., Devoret, M. H., and Schoelkopf, R. J. (2014). Tracking photon jumps with repeated quantum non-demolition parity measurements. *Nature*, 511(7510):444–448. (Cited on pages 61, 67, and 70).

[Tao et al., 2019] Tao, X., Chen, S., Chen, Y., Wang, L., Li, X., Tu, X., Jia, X., Zhao, Q., Zhang, L., Kang, L., and Wu, P. (2019). A high speed and high efficiency superconducting photon number resolving detector. *Superconductor Science and Technology*, 32(6):064002. (Cited on page 60).

[Temme et al., 2017] Temme, K., Bravyi, S., and Gambetta, J. M. (2017). Error Mitigation for Short-Depth Quantum Circuits. *Physical Review Letters*, 119(18):180509. (Cited on page 75).

[Teoh, 2023] Teoh, J. D. (2023). *Error Detection in Bosonic Circuit Quantum Electrodynamics*. PhD thesis, Yale University. (Cited on pages 89 and 90).

[Teoh et al., 2023] Teoh, J. D., Winkel, P., Babla, H. K., Chapman, B. J., Claes, J., de Graaf, S. J., Garmon, J. W. O., Kalfus, W. D., Lu, Y., Maiti, A., Sahay, K., Thakur, N., Tsunoda, T., Xue, S. H., Frunzio, L., Girvin, S. M., Puri, S., and Schoelkopf, R. J. (2023). Dual-rail encoding with superconducting cavities. *Proceedings of the National Academy of Sciences*, 120(41):e2221736120. (Cited on pages 11, 13, 89, 90, 108, and 182).

[Tillmann et al., 2013] Tillmann, M., Dakić, B., Heilmann, R., Nolte, S., Szameit, A., and Walther, P. (2013). Experimental boson sampling. *Nature Photonics*, 7(7):540–544. (Cited on page 60).

[Tsunoda et al., 2023] Tsunoda, T., Teoh, J. D., Kalfus, W. D., de Graaf, S. J., Chapman, B. J., Curtis, J. C., Thakur, N., Girvin, S. M., and Schoelkopf, R. J. (2023). Error-Detectable Bosonic Entangling Gates with a Noisy Ancilla. *PRX Quantum*, 4(2):020354. (Cited on pages 114, 128, 129, 130, 168, 178, and 182).

[Unanyan et al., 1998] Unanyan, R., Fleischhauer, M., Shore, B. W., and Bergmann, K. (1998). Robust creation and phase-sensitive probing of superposition states via stimulated Raman adiabatic passage (STIRAP) with degenerate dark states. *Optics Communications*, 155(1):144–154. (Cited on page 134).

[Ursin et al., 2007] Ursin, R., Tiefenbacher, F., Schmitt-Manderbach, T., Weier, H., Scheidl, T., Lindenthal, M., Blauensteiner, B., Jennewein, T., Perdigues, J., Trojek, P., Ömer, B., Fürst, M., Meyenburg, M., Rarity, J., Sodnik, Z., Barbieri, C., Weinfurter, H., and Zeilinger, A. (2007). Entanglement-based quantum communication over 144 km. *Nature Physics*, 3(7):481–486. (Cited on page 59).

[van Merriënboer et al., 2019] van Merriënboer, B., Breuleux, O., Bergeron, A., and Lamblin, P. (2019). Automatic differentiation in ML: Where we are and where we should be going. arXiv:1810.11530 [cs, stat]. (Cited on page 123).

[Venkatraman et al., 2023] Venkatraman, J., Cortinas, R. G., Frattini, N. E., Xiao, X., and Devoret, M. H. (2023). A driven quantum superconducting circuit with multiple tunable degeneracies. arXiv:2211.04605 [cond-mat, physics:physics, physics:quant-ph]. (Cited on pages 10 and 38).

[Vittal et al., 2023] Vittal, S., Das, P., and Qureshi, M. (2023). ERASER: Towards Adaptive Leakage Suppression for Fault-Tolerant Quantum Computing. (Cited on page 132).

[Wang, 2022] Wang, C. S. (2022). *Bosonic Quantum Simulation in Circuit Quantum Electrodynamics*. PhD thesis, Yale University. (Cited on page 64).

[Wang et al., 2020] Wang, C. S., Curtis, J. C., Lester, B. J., Zhang, Y., Gao, Y. Y., Freeze, J., Batista, V. S., Vaccaro, P. H., Chuang, I. L., Frunzio, L., Jiang, L., Girvin, S., and Schoelkopf, R. J. (2020). Efficient Multiphoton Sampling of Molecular Vibronic Spectra on a Superconducting Bosonic Processor. *Physical Review X*, 10(2):021060. (Cited on pages 11, 13, 21, 24, 38, 59, 62, 64, 65, and 85).

[Wang et al., 2023] Wang, C. S., Frattini, N. E., Chapman, B. J., Puri, S., Girvin, S., Devoret, M. H., and Schoelkopf, R. J. (2023). Observation of Wave-Packet Branching through an Engineered Conical Intersection. *Physical Review X*, 13(1):011008. (Cited on page 11).

[Whitfield et al., 2011] Whitfield, J. D., Biamonte, J., and Aspuru-Guzik, A. (2011). Simulation of electronic structure Hamiltonians using quantum computers. *Molecular Physics*, 109(5):735–750. (Cited on page 63).

[Wineland et al., 1978] Wineland, D. J., Drullinger, R. E., and Walls, F. L. (1978). Radiation-Pressure Cooling of Bound Resonant Absorbers. *Physical Review Letters*, 40(25):1639–1642. (Cited on page 8).

[Wiseman and Milburn, 2014] Wiseman, H. M. and Milburn, G. J. (2014). *Quantum measurement and control*. Cambridge University Press, Cambridge. (Cited on page 123).

[Wood and Gambetta, 2018] Wood, C. J. and Gambetta, J. M. (2018). Quantification and characterization of leakage errors. *Physical Review A*, 97(3):032306. (Cited on pages 143 and 144).

[Wu et al., 2022] Wu, Y., Kolkowitz, S., Puri, S., and Thompson, J. D. (2022). Erasure conversion for fault-tolerant quantum computing in alkaline earth Rydberg atom arrays. *Nature Communications*, 13(1):4657. (Cited on pages 11, 13, and 89).

[Wölk et al., 2015] Wölk, S., Piltz, C., Sriarunothai, T., and Wunderlich, C. (2015). State selective detection of hyperfine qubits. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 48(7):075101. (Cited on page 76).

[Xu et al., 2023] Xu, Q., Zeng, P., Xu, D., and Jiang, L. (2023). Fault-Tolerant Operation of Bosonic Qubits with Discrete-Variable Ancillae. (Cited on page 181).

[Yin et al., 2016] Yin, H.-L., Chen, T.-Y., Yu, Z.-W., Liu, H., You, L.-X., Zhou, Y.-H., Chen, S.-J., Mao, Y., Huang, M.-Q., Zhang, W.-J., Chen, H., Li, M. J., Nolan, D., Zhou, F., Jiang, X., Wang, Z., Zhang, Q., Wang, X.-B., and Pan, J.-W.

(2016). Measurement-Device-Independent Quantum Key Distribution Over a 404 km Optical Fiber. *Physical Review Letters*, 117(19):190501. (Cited on page 59).

[Zhang et al., 2017] Zhang, W., Ding, D.-S., Sheng, Y.-B., Zhou, L., Shi, B.-S., and Guo, G.-C. (2017). Quantum Secure Direct Communication with Quantum Memory. *Physical Review Letters*, 118(22):220501. (Cited on page 59).

[Zhang et al., 2022] Zhang, Y., Curtis, J. C., Wang, C. S., Schoelkopf, R. J., and Girvin, S. M. (2022). Drive-induced nonlinearities of cavity modes coupled to a transmon ancilla. *Physical Review A*, 105(2):022423. (Cited on pages 12, 17, and 53).

[Zhang et al., 2019] Zhang, Y., Lester, B. J., Gao, Y. Y., Jiang, L., Schoelkopf, R. J., and Girvin, S. M. (2019). Engineering bilinear mode coupling in circuit QED: Theory and experiment. *Physical Review A*, 99(1):012314. (Cited on page 56).

[Zheng et al., 2022] Zheng, W., Zhang, Y., Dong, Y., Xu, J., Wang, Z., Wang, X., Li, Y., Lan, D., Zhao, J., Li, S., Tan, X., and Yu, Y. (2022). Optimal control of stimulated Raman adiabatic passage in a superconducting qudit | npj Quantum Information. *npj Quantum Information*, 8(1):9. (Cited on page 134).